现代应用数学丛书

群体遗传学的数学理论

（日）木村资生 著

上海科学技术出版社

现代应用数学丛书

群体遺傳学的数学理论

〔日〕木村資生 著
刘 祖 洞 譯

上海科学技术出版社

內 容 提 要

本书是日本岩波书店出版的現代应用数学丛书之一的中譯本。扼要介紹群体遺傳学的数学理論，以自然群体为主，广泛应用了概率論和数理統計的理論和方法，建立了許多重要的概念和公式。全书共分四章，第1章介紹孟德尔群体的基本性质，第2章叙述自然选擇的作用，第3章介紹突变的效应，最后一章叙述随机过程理論在群体遺傳学中的应用。可供生物科学工作者和数学工作者作为参考。

現代应用数学丛书

群体遺傳学的数学理论

原 书 名　集團遺伝学の数学的理論
原 著 者　〔日〕木 村 資 生
原出版者　岩 波 书 店
譯　　者　刘 祖 洞

上海科学技术出版社出版
（上海瑞金二路450号）
上海市书刊出版业营业許可証出093号
新华书店上海发行所发行　各地新华书店經销
商务印书館上海厂印刷

开本850×1168 1/32　印張2 2/32　字数47,000
1962年5月第1版　1962年5月第1次印刷
印数 1—4,000

統一书号：13119・458
定　价：（十四）0.38元

出 版 說 明

这一套书是根据日本岩波书店出版的"現代应用数学讲座"翻譯而成。日文原书共15卷60册，分成A、B两組，各編有序号。現在把原来同一题目分成两册或三册的加以合并，整理成42种，不另分組編号，陆續翻譯出版。

这套书涉及的面很广，其内容都和現代科学技术密切有关，有一定参考价値。每一本书收集的資料都比較丰富，而叙述扼要，篇幅不多，有利于讀者以較短时間掌握有关学科的主要內容。虽然，这套书的某些观点不尽适合于我国的情况，但其方法可供参考。因此，翻譯出版这一套书，对我国学术界是有所助益的。

由于日文原书是1957年起以讲座形式陆續出版的，写作时間和篇幅的限制不可避免地会影响原作者对内容的处理，为了尽可能地减少这种影响，我們在每一譯本中，特請譯者或校閱者撰写序或后記，以介紹有关学科的最近发展状况，并对全书内容作一些評价，提出一些看法，結合我国情况补充一些資料文献，在文内过于簡略或不足的地方添加了必要的注释和改正原书中存在的一些錯誤。希望这些工作能对讀者有所帮助。

承担翻譯和校閱的同志，为提高书籍的质量付出了巨大劳动，在此特致以誠摯的謝意。

欢迎讀者对本书提出批評和意見。

上海科学技术出版社

譯 者 序

孟德尔的遺傳法則是說明杂交中基因的行为的。基因的行为条理井然，有規可循，具有內在的数学美（mathematical beauty)。所以以孟德尔法則为基础而发展起来的遺傳学，从它誕生之日起，就和数学有着很密切的关系。下面几个遺傳学分支就是运用了数学的理論和方法，逐漸丰富或建立起来的。

(1) 群体遺傳学方面：达尔文已认識到，生物体的微細的变异在进化上是很重要的，可作进化的原材料，供自然选擇之用。可是有同一变异的个体数少，不能不与沒有这种变异的个体交配。所以每經过一次交配，这种变异的强度就减弱一次。为了提供充分的、可供选擇的材料，非有大量的微細变异經常地发生不可，而这又是和事实不相符合的。所以这是达尔文无法解决的难題。直到1908年，Hardy 和 Weinberg 各別地指出，根据孟德尔的分离法則，用数学方法来推算，在一个很大的、随机婚配的群体中，新发生的突变是可以一代代地維持下来，它們的强度不会减弱，因而通过自然选擇的作用，是可以作为进化的原材料的。这样，达尔文无法解决的难題解决了。以后又由数学家和遺傳学家进而从群体的角度，对进化的各要素——突变、选擇、隔离和迁移等过程作了縝密的数学的探討，而本书作者也对基因頻率在群体中随机变动等問題作出了一定的业績，使我們对进化过程的认識有所提高，卒能使运用数学方法来研究进化問題的群体遺傳学成为遺傳学中的一个重要分支。

(2) 人类遺傳学方面：自从孟德尔法則被重新发現后，不久就在人类性状、特別是病理性状的遺傳方面得到証实。不过研究

人类的遺傳，不能随意地进行实驗，所以除了进行临床观察外，还得进行家系的分析。可是人类每个家庭中子女数少，为了正确地决定遺傳方式，还須从群体的角度，进行統計学的分析。例如 ABO 血型的遺傳方式，以前以为是由二对基因决定的，經过統計的分析后，知道是由同一位点 (locus) 上的三个复等位基因 (multiple alleles) 决定的。此外，由于近年来輻射能的广泛应用，人們接触射綫的机会相应地增加，这样自然会提高人类的突变頻率，所以我們要知道是否存在有"可允許的輻射剂量"，假使有的話，最低的安全剂量是多少，这些問題都是要通过数学的分析，才可得出某种答案的。所以人类遺傳学中的很多发展都是和数学的应用有着直接关系的。

(3) 数量遺傳学方面：遺傳学上的研究，大多偏重于单因子的遺傳。因为它們的作用显著，分离清楚，受环境的影响小，所以可以用比較簡单的方式来进行分析。可是經济动植物的很多生产性能，例如乳牛的泌奶量，雌鸡的产卵率等，往往是多因子遺傳的。这些因子的显性常常不完全，个别基因的效应很小，有着累加的作用，而且受环境的影响很大，所以一定要用統計数学的方法来进行研究。近年来，这方面的研究，进展非常快，运用方差分析等各种方法，計算遺傳力 (heretability) 和环境的作用，使进行动植物育种工作时有所依据，育种效果的預見性也加强了。所以数量遺傳学旣有丰富的理論，而又密切联系实际，是近年来遺傳学中迅速发展的生长点之一。其所以能够如此，数学理論的运用显然是起着决定性作用的。

从上面看来，数学和遺傳学的关系是紧紧地結合在一起的，遺傳学中有些分支的发展是和数学理論，尤其是統計理論的应用分不开的。所以为了使遺傳学有更快的进展，一方面要遺傳学家加强数学方面的修养，能从数学的角度来思考遺傳学問題；而另一方

面；也要引起数学家对遺傳学的兴趣，来丰富遺傳学中的数学理論。木村資生著的这本书虽然只着重在群体遺傳学方面，对人类遺傳学与数量遺傳学的数学理論只略略提及，但是通过这本薄薄的小册子，如能使我国数学家和遺傳学家了解数学理論在遺傳学研究中的作用，引起大家的兴趣，进行这方面的研究，那末这就是譯者的最大欣慰了。

刘祖洞 1961 年 12 月

序

群体遺傳学的数学理論开始成为一門有系統的学問，是1930年左右 Fisher, Wright 和 Haldane 三人所进行的划时代研究的結果。有了这些工作，Darwin 用自然选擇来說明的进化学說，就通过概率論和統計学而与 Mendel 最初发現的遺傳法則綜合起来了。在这个意义上，他們的功績在近代生物学史上是应該特別記上一笔的。以后在这方面的研究，除了上述三人以外，由于遺傳学者、数学者、統計学者、动植物育种学者的研究，各方面都有进展，不仅在自然群体的研究、种的分化問題方面，而且在动植物的科学育种法、人类遺傳学、优生学，进而在放射綫影响人类群体的伤害問題等方面，数学理論都成为不可缺少的了。要以有限的篇幅全面地說明这些問題，毕竟是不可能的。所以下面主要着眼于自然群体，叙述基本理論，但在个别地方，把重要的公式或概念，納入练习題中，尽可能地使本书的內容丰富。

目　录

第1章　Mendel 群体的基本性质

在这一章中，研究突变、自然选择等不起作用时，单是由于 Mendel 的机构，群体是在怎样的平衡状态上移动的。

§1 Hardy-Weinberg 法则

在常染色体上取一个基因位点，为简单起見，先考虑有二个等位基因的例子[1]。假定群体很大，雄和雌的三种遺傳型 AA, AA'，$A'A'$ 的相对频率(以下簡称频率)各为 P_m, Q_m, R_m 与 P_f, Q_f, R_f，

$$P_m+Q_m+R_m=P_f+Q_f+R_f=1.$$

故如基因 A 和 A' 在雄的频率为 p_m, q_m，在雌的频率为 p_f, q_f，則

$$p_m=P_m+\frac{1}{2}Q_m, \quad q_m=1-p_m,$$

$$p_f=P_f+\frac{1}{2}Q_f, \quad q_f=1-p_f.$$

所以移到下一代时，如交配完全随机(random mating)，則各遺傳型的交配频率就是下式的展开:

$$(AAP_m+AA'Q_m+A'A'R_m)(AAP_f+AA'Q_f+A'A'R_f).$$

例如 AA 雄×AA 雌的频率是 P_mP_f. 由这些交配，得到次代的各遺傳型，它們的频率如表1.1所示。如次代群体中三遺傳型 AA，AA', $A'A'$ 的频率各为 P, Q, R(雌雄相同)，則

---

[1] 数学者、統計学者想得到遺傳学的知識，可参閱方宗熙著《細胞遺傳学》，科学出版社；北京农业大学編《普通遺傳学》，农业出版社；上海复旦大学遺傳研究室編的《摩尔根学派的遺傳学說》，北京科学普及出版社；以及本书卷末所推荐的 Sturtevant & Beadle 著 "An Introduction to Genetics"，美国 Philadelphia 的 Saunders 图书公司出版。——譯者注

$$P=P_mP_f+\frac{1}{2}(P_mQ_f+Q_mP_f)+\frac{1}{4}Q_mQ_f$$
$$=\left(P_m+\frac{1}{2}Q_m\right)\left(P_f+\frac{1}{2}Q_f\right)=p_mp_f.$$

同样

$$Q=p_mq_f+q_mp_f,\quad R=q_mq_f,$$

从而

$$P+Q+R=1.$$

以上的結果等于

$$(Ap_m+A'q_m)(Ap_f+A'q_f)$$

的展开 $AAp_mp_f+\cdots$，这一点象下面那样来考虑就容易理解。先考虑雄的方面，含基因 A 的精子的頻率是 p_m，含基因 A' 的精子的頻率是 q_m. 其次，在雌的方面，含 A 和 A' 的卵子的頻率各为 p_f 和 q_f，所以在完全随机的交配下，A 精子使 A 卵子受精的概率是 p_mp_f，由此产生次代的 AA。同样，A 精子使 A' 卵子受精，以及 A' 精子使 A 卵子受精的頻率各为 p_mq_f 和 q_mp_f，由此产生 AA'。又 $A'A'$ 由 A' 精子与 A' 卵子結合而生，它的頻率是 q_mq_f。这里所考虑的基因是在常染色体上的，与决定性别的基因是独立地行

表 1.1

交配型式	頻率	子代的分离比
$AA\times AA$	P_mP_f	$1AA$
$AA\times AA'$ $AA'\times AA$	P_mQ_f Q_mP_f	$\frac{1}{2}AA:\frac{1}{2}AA'$
$AA\times A'A'$ $A'A'\times AA$	P_mR_f R_mP_f	$1AA'$
$AA'\times AA'$	Q_mQ_f	$\frac{1}{4}AA:\frac{1}{2}AA':\frac{1}{4}A'A'$
$AA'\times A'A'$ $A'A'\times AA'$	Q_mR_f R_mQ_f	$\frac{1}{2}AA':\frac{1}{2}A'A'$
$A'A'\times A'A'$	R_mR_f	$1A'A'$

动的，所以三遗傳型的頻率 P, Q, R 对雌雄是一样的，如基因 A 和 A' 的頻率各为 p 和 $q(p+q=1)$，則

$$p=P+\frac{1}{2}Q=p_mp_f+\frac{1}{2}(p_mq_f+q_mp_f)$$
$$=\frac{1}{2}p_m+\frac{1}{2}p_f.$$

同样

$$q=\frac{1}{2}q_m+\frac{1}{2}q_f.$$

卽雌雄間基因頻率不同时，在下一代就平均起来，两性变为一样。所以在此后的世代中，可以看作 $p_m=p_f$ 的特例，导出下面的結論。

交配完全随机，选擇、突变、移住不起作用，則基因頻率及三种遺傳型的頻率每代不变，如把配子中的基因頻率系列 (gametic array) 写作

$$pA+qA',$$

則接合子中遺傳型頻率系列(zygotic array)是

$$p^2AA+2pqAA'+q^2A'A', \tag{1.1}$$

換句話說，可用前者的平方来表示。这結果是英国数学者 Hardy 和德国医生 Weinberg 在1908年各自发表的，所以附有两人的名字，就是現在周知的 Hardy-Weinberg 法則。对历史背景有兴趣的人，可参閱卷末所列参考文献中 Stern.(1943) 的报告。把上面的結果推广，設群体中任意个(n)等位基因 A_1, A_2, …, A_n 的頻率是 p_1, p_2, …, p_n $\left(\sum_1^n p_i=1\right)$，可以証明，遺傳型頻率系列能表为

$$(p_1A_1+p_2A_2+\cdots+p_nA_n)^2. \tag{1.2}$$

习題 ABO 血型是由三个等位基因支配的。为方便計，把它們記作 A, B, O，則 A 和 B 对 O 都是显性，AA 和 AO 个体是 A 血型，BB 和 BO 个体是 B 型，AB 和 OO 各为 AB 型和 O 型。以 A, B, O 三基因的頻率为 p, q, r，則日本人大約 $p=0.30$, $q=0.16$, $r=0.54$，由此算出各血型的頻率。

§2 伴性遺傳

考虑性染色体 X 上的位点(locus)，二个等位基因为 A 和 A'。Y 染色体是空虚的，所以写作 O，这样雄可以有 AO 和 $A'O$ 两型，而雌可以有 AA，AA' 和 $A'A'$ 三型。现在，如这些遺傳型在第 n 代时的相对頻率各为 S_n，T_n $(S_n+T_n=1)$ 及 P_n，Q_n，R_n $(P_n+Q_n+R_n=1)$，則由完全随机的交配，次代各遺傳型的頻率，雄的是

$$O(p_nA+q_nA')$$

的展开，而雌的是

$$(S_nA+T_nA')(p_nA+q_nA')$$

的展开。这里，p_n 和 q_n 是第 n 代的雌的 A 和 A' 的頻率:

$$p_n=P_n+\frac{1}{2}Q_n,\quad q_n=\frac{1}{2}Q_n+R_n,\quad p_n+q_n=1.$$

所以次代各型的頻率是

$$S_{n+1}=p_n,\quad T_{n+1}=q_n,$$

$$P_{n+1}=S_np_n,\quad Q_{n+1}=S_nq_n+T_np_n,\quad R_{n+1}=T_nq_n.\qquad(2.1)$$

因此雄的 AO 和 $A'O$ 的頻率各各等于前代 A 和 A' 的頻率，又

$$\begin{aligned}p_{n-1}-p_n&=\left(P_{n+1}+\frac{1}{2}Q_{n+1}\right)-p_n\\&=S_np_n+\frac{1}{2}(S_nq_n+T_np_n)-p_n\\&=S_np_n+\frac{1}{2}S_nq_n+\frac{1}{2}(1-S_n)p_n-p_n\\&=\frac{1}{2}(S_n-p_n)=-\frac{1}{2}(p_n-p_{n-1})\\&=\left(-\frac{1}{2}\right)^n(p_1-p_0)\quad(n=1,2,3,\cdots),\end{aligned}$$

从而可以知道，雌的基因頻率在相继世代間的差异，每一代减少一半。所以經过相当世代后，基因頻率接近于某极限值，如令这极限

值分别为 p_∞ 和 q_∞，則

$$S_\infty = p_\infty,\quad T_\infty = q_\infty,$$

$$P_\infty = p_\infty^2,\quad Q_\infty = 2p_\infty q_\infty,\quad R_\infty = q_\infty^2,$$

从而雄、雌的基因頻率就相等了。

习題 1 証明第 n 代的雌的基因頻率 p_n 是

$$p_n = p_0 + \frac{2}{3}\left[1-\left(-\frac{1}{2}\right)^n\right](p_1 - p_0),$$

用图表示基因頻率随世代变化的情形。

习題 2 从 $AO\times AA'$ 的交配出发，各代的基因頻率是怎样变化的？

§3 連鎖的效应

考虑在一个常染色体上的二个基因位点，第一个位点存在有 n 个复等位基因 $A_1, A_2, \cdots, A_n$，它們的頻率是 $x_1, x_2, \cdots, x_n$，第二个位点存在有 m 个基因 $B_1, B_2, \cdots, B_m$，它們的頻率是 $y_1, y_2, \cdots, y_m$。設 A_iB_k 的染色体在第 t 代时的頻率为 $P_{ik}^{(t)}$，两基因位点間的交换率(recombination fraction)为 r。現在，如交配完全随机，考虑第 t 代时的 A_iB_k 的染色体，則这或者是从上代的 $A_iB_k/--$ 个体的未經交叉的染色体来的，或者是从上代的 $A_i-/-B_k$ 个体的經过交叉的染色体来的。这里横划(—)表示任意的基因。所以

$$P_{ik}^{(t)} = (1-r)P_{ik}^{(t-1)} + rx_iy_k. \tag{3.1}$$

这里

$$P_{ik}^{(t)} = x_iy_k + \Delta_{ik}^{(t)},$$

代入上式的两边，得

$$\Delta_{ik}^{(t)} = (1-r)\Delta_{ik}^{(t-1)},$$

所以

$$\Delta_{ik}^{(t)} = (1-r)^t\Delta_{ik}^{(0)},$$

$$P_{ik}^{(t)} = x_iy_k + (1-r)^t\Delta_{ik}^{(0)}. \tag{3.2}$$

这里，r 取 0 与 1/2 間的值，所以只要不是 $r=0$，就有

$$\lim_{t\to\infty} \Delta_{ik}^{(t)}=0, \quad \lim_{t\to\infty} P_{ik}^{(t)}=x_i y_k.$$

即經过相当代数后，染色体 A_iB_k 的頻率等于基因 A_i 的頻率与基因 B_k 的頻率的乘积。所以群体在随机交配下，經过平衡后，对于两性状間的相关，普通不能认为是由于基因間的連鎖的緣故。

习题 1 就第一和第二位点上各有二个等位基因 A, A' 和 B, B' 时，直接驗証(3.1)。

习题 2 就同样的場合，若初期世代只有 $AB/A'B'$ 的个体，交换值是1% ($r=0.01$)时，离开最后可以到达的平衡状态相差不到1%，大概要經过多少代数?

习题 3 雄、雌的交换值不等时，(3.1)必須作怎样的修正?

§4 內交配(近亲交配)

很久以来，許多学者进行了这方面的研究，但主要的結果大部分都与 Wright, Haldane, Fisher 和 Malécot 的工作有关。在这里，以自然群体研究中必需的內交系数(inbreeding coefficient) F 为中心来加以說明。根据 Malécot (1948)，F 是这样地导入的。取任一个体，設它所有的二个相同基因从祖先的某一基因而来的概率为 F. 所以同一个体中的二个相同基因以概率 $1-F$ 从互不相关的祖先分别地得来。这里，設第 i 个等位基因 A_i 的相对頻率是 x_i，同型接合体 A_iA_i 的頻率为 P_{ii}，异型接合体 $A_iA_j(i\neq j)$ 的頻率为 $2P_{ij}$. 或者为了使定义簡化，在异型接合体时，把半数写作 A_iA_j，而把其余半数写作 A_jA_i，这样，同型、异型接合体就沒有区别，A_iA_j 的頻率也可写作 P_{ij}. 不消說，$P_{ij}=P_{ji}$. 在遺傳学上，这样的区别一般沒有意义，但进行計算时非常便利。以下暫且把它叫做**順序性遺傳型**。

这里，我們来探討一下任意取出的个体正好是同型接合体 A_iA_i 的概率。二个基因来自同一祖先时，两者始終有同一的記号，

两者均为 A_i 的概率是 x_i。其次，二个基因来自不同祖先时，从独立事件的計算法可知，两者均为 A_i 的概率是 x_i^2。从而

$$P_{ii}=Fx_i+(1-F)x_i^2.$$

至于任意取出的个体恰为异型接合体的概率，因为二个基因不可能来自共同的祖先基因，所以是

$$P_{ij}=(1-F)x_ix_j.$$

当等位基因只有二个时，設这二个基因为 A 和 A'，并設三个遺傳型 AA, AA', $A'A'$ 的频率是 P, $2Q$, R, 則

$$P=Fx+(1-F)x^2,$$

$$Q=(1-F)x(1-x),$$

$$R=F(1-x)+(1-F)(1-x)^2.$$

这里 x 是 A 的频率。以上的結果写成表 4.1，把这表看作相关表，設 A 的值为 1, A' 的值为 0, 計算相关系数，得

$$r=\frac{P-x^2}{x(1-x)}=\frac{Fx+(1-F)x^2-x^2}{x(1-x)}=F.$$

表 4.1

		(1) A	(0) A'	
(1)	A	P	Q	x
(0)	A'	Q	R	$1-x$
		x	$1-x$	

事实上，Wright (1921) 把 F 定义为“結合的配子間的相关”。交配完全随机时，$F=0$; 长年間继續强度的近亲交配，一切个体都成同型接合体状态，此时 $F=1$。

其次来探討一下群体遺傳学上的一个重要情况，就是一个群体被分为几个分群体 (subgroup)。設等位基因中的一个特定基

因为 A，其余的概括起来記作 A'。設第 i 个分群体的相对大小为 $w_i(\sum_i w_i=1)$，其中 A 的频率为 x_i。全群体中 A 的频率 x 就是

$$x=\sum_i w_i x_i.$$

現在設任意个体对全群体的內交系数为 F_{IT}，对它所属的分群体的內交系数为 F_{IS}。AA' 在全群体中的频率等于

$$2(1-F_{IT})x(1-x);$$

这里，如用 P 来代替 $1-F$，依据 Wright(1951)，把 P 称为完全混合系数(panmictic index)，則

$$\begin{aligned} P_{IT}x(1-x) &= P_{IS}\sum_i x_i(1-x_i)w_i \\ &= P_{IS}\{x(1-x)-V_x\}. \qquad (4.1) \end{aligned}$$

这里，V_x 是分群体間的基因频率的方差，即

$$V_x=\sum_i (x_i-x)^2 w_i=\sum_i x_i^2 w_i-x^2.$$

其次，从各分群体随机地抽取二个基因，設其在全群体的相关系数为 F_{ST}，則

$$F_{ST}=V_x/x(1-x). \qquad (4.2)$$

要求出这个值，可这样地考虑。从第 i 个分群体独立而随机地抽取的二个基因記为 X_i 和 Y_i，分别地，如抽取的基因是 A，給与值 1，如抽取的基因是 A'，則給与值 0。X_i 在分群体的均值为 x_i，在全群体的均值为 x，同样有 Y_i 的均值。如以 E 表求取数学期望的运算符号，則 X_i 和 Y_i 在全群体中的协方差是

$$\mathrm{E}\sum_i w_i(X_i-x)(Y_i-x)=\sum_i w_i(x_i-x)(x_i-x)=V_x.$$

又 X_i 和 Y_i 在全群体的方差都是 $x(1-x)$。F_{ST} 可以认为是分群体对全群体的內交系数。以 P_{ST} 代替 $1-F_{ST}$，則由(4.2)得

$$P_{ST}=\frac{x(1-x)-V_x}{x(1-x)},$$

从而与(4.1)合并，得到富有意义的关系式

$$P_{IT}=P_{IS}P_{ST}. \tag{4.3}$$

換言之，个体对全群体的完全混合系数，等于个体对分群体的完全混合系数与分群体对全群体的完全混合系数之积。同样，各分群体(S)再分为小分群体(S')时，可以証明

$$P_{IT}=P_{IS'}P_{S'S}P_{ST}.$$

习题1 考慮一群体，交配完全随机，設异型接合体 AA' 的频率为 H_0。其次，分这群体为分群体，各自进行完全随机交配时，設那时的异型接合体的频率为 H_1，則

$$H_1=H_0-2V_x,$$

这里 V_x 是分群体間基因频率的方差。这事实已由 Wahlund (1928) 指出(Wright, 1931)。

习题2 在部分杂交的群体中，設自株授精率为 μ，异株授精率为 λ($\mu+\lambda=1$)，第 t 代异型接合体 A_iA_j 的频率为 $H_{ij}^{(t)}$，則

$$H_{ij}^{(t+1)}=2\lambda x_ix_j+\frac{\mu}{2}H_{ij}^{(t)} \quad (i\neq j),$$

这里，x_i 是第 i 个等位基因 A_i 的频率。从而第 t 代的 A_iA_j 的频率是

$$H_{ij}^{(t)}=H_{ij}^{(\infty)}+\left(\frac{\mu}{2}\right)^t\left[H_{ij}^{(0)}-H_{ij}^{(\infty)}\right],$$

經过相当数目的世代后，証明可把它归結为

$$H_{ij}^{(\infty)}=4\lambda x_ix_j/(2\lambda+\mu).$$

在这个平衡状态时，同型接合体 A_iA_i 的频率是

$$H_{ii}^{(\infty)}=(\mu x_i+2\lambda x_i^2)/(2\lambda+\mu).$$

§5 近緣系数

以下如沒有預先声明，探討的都是一个基因位点，但群体中可以存在有任何数目的等位基因。現在設群体中特定的二个体为 I 和 J，I 所有的一对相同基因中的任何一个与 J 所有的任何一个，来自祖先的同一基因的概率称为 I 与 J 間的近緣系数(coefficient of relationship)，用 F_{IJ} 表示。上一节中說明的內交系数，是同一个体內的相同基因間的近緣系数。个体間的系譜知道时，可根据

它算出二个体 I, J 間的近緣系数。若两者間沒有共同祖先，不消說，近緣系数的值是 0. 其次，如图 5.1 所示，設两者間仅有一个共同祖先 A，由 I 追溯上去，相当于第 m 代，由 J 追溯上去，相当于第 n 代。我們来探討一下通过 A 而把 I 和 J 連結起来的通路。由 I 取得的任意

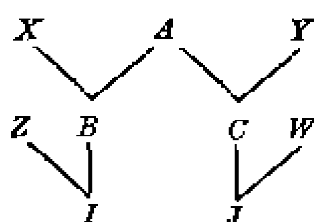


图　5.1

基因 α，来自这个通路上前一代祖先(图中 B)的概率是 1/2，来自 m 代前的祖先 A 的概率是 $(1/2)^m$. 同样，由 J 取得的任意基因 β (与 α 相同)，来自 n 代前的祖先 A 的概率是 $(1/2)^n$. 所以 α 与 β 来自同一个体 A 的概率是 $(1/2)^{m+n}$. 其中半数的場合，α 与 β 是由 A 内的同一基因来的，而其余的場合，是由 A 内各别的相同基因来的。在后者的場合，这个相同基因又是从祖先的共同基因来的，它的概率等于 F_A，也就是等于 A 的内交系数，所以把它們綜合起来，得到

$$F_{IJ}=\left(\frac{1}{2}\right)^{m+n}\frac{1+F_A}{2}.$$

如图 5.1 的例子，$F_A=0$，所以

$$F_{IJ}=\left(\frac{1}{2}\right)^{2+2}\frac{1+0}{2}=\left(\frac{1}{2}\right)^5.$$

其次，I 与 J 間有几个共同祖先 A_1, A_2, $\cdots$ 时，設第 i 个祖先为 A_i，相当于 A_i 的 m 和 n 各为 m_i 和 n_i，則一般式是

$$F_{IJ}=\sum_i\left(\frac{1}{2}\right)^{m_i+n_i}\frac{1+F_{Ai}}{2}. \tag{5.1}$$

习題 1　就图 5.2～5.5 所示的四个例子(兄弟，亲子，半兄弟，表兄弟)，計算 F_{IJ}. 但 $F_A=F_B=0$.

习題 2　如图 5.6 所示，继續兄妹交配时，如第 n 代的个体的内交系数为 F_n，各别二个体間的近緣系数为 ϕ_n，証明下面的关系式成立:

$$\phi_n=\frac{1}{2}\phi_{n-1}+2\left(\frac{1}{2}\right)^2\frac{1+F_{n-1}}{2},\qquad F_n=\phi_{n-1}.$$

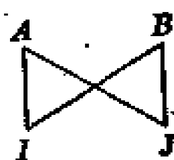


图 5.2

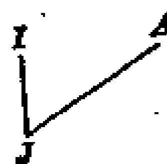


图 5.3

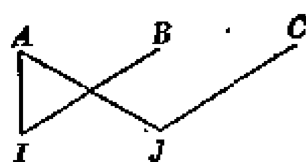


图 5.4

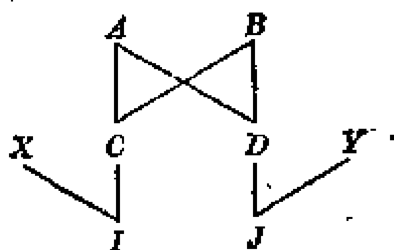


图 5.5

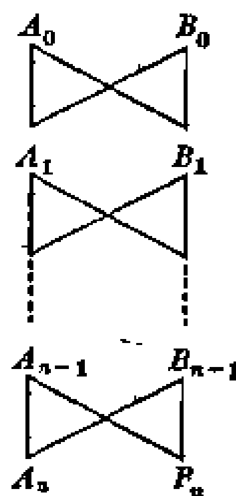


图 5.6

由此，使 $P_n=1-F_n$，则得差分方程：

$$P_{n+1}=\frac{P_n}{2}+\frac{P_{n-1}}{4}. \tag{5.2}$$

这里，P_n 是在第 n 代的完全混合系数，表示一个体中的相同基因来自独立无关的祖先的概率。如前节所示，异型接合体的频率 H_n 与 P_n 成比例，试証

$$H_n=A\varepsilon^n+A'\varepsilon'^n,$$

这里
$$A=\frac{2}{\sqrt{5}}(H_1-\varepsilon' H_0),\quad A'=\frac{2}{\sqrt{5}}(H_0\varepsilon-H_1),$$

$$\varepsilon=\frac{1+\sqrt{5}}{4},\quad \varepsilon'=\frac{1-\sqrt{5}}{4}.$$

§6 数量性状的变差分析

Fisher(1918)，Wright(1935)等的研究，奠定了这方面的基

础。近年来，Kempthorne(1956) 的貢献是很突出的。先来考虑一个基因位点，存在有任意个等位基因 $A_1, A_2, \cdots, A_n$。遺傳型 A_iA_j 在特定环境下的数量性状的值，称为它的表現型值 (phenotypic value)，这个值在可能的全环境中的平均值，称为遺傳型值 (genotypic value)。以下主要着眼于遺傳型值的分析。現在来考虑一下，設 A_iA_j 的遺傳型值为 $Y_{ij}(Y_{ij}=Y_{ji})$，它可表示为以下的形式：

$$Y_{ij}=\overline{Y}+\alpha_i+\alpha_j+d_{ij} \tag{6.1}$$

这里，$\overline{Y}$ 是遺傳型值的群体平均，α_i 和 α_j 各为基因 A_i 和 A_j 的加算效应，选擇而使

$$Q=\sum_{ij}P_{ij}(Y_{ij}-\overline{Y}-\alpha_i-\alpha_j)^2 \tag{6.2}$$

的值成为最小。其中，P_{ij} 是順序性遺傳型 A_iA_j 的相对頻率 $(P_{ij}=P_{ji})$．所以，应用綫性形式(6.1)时，求得的 α_i 就是基因 A_i 效应的最可能值。換句話說，就是从遺傳型值对基因数的回归直綫

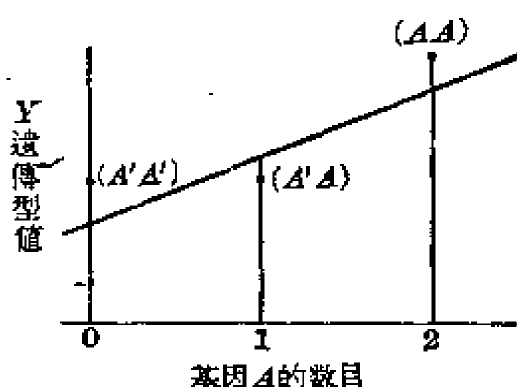


图 6.1

$$Y_{ij}=\overline{Y}+\alpha_i+\alpha_j$$

来看时，α_i 就是它的回归系数。图6.1表示等位基因只有 A 和 A' 二个时的情况。为了决定 α 值，使 $1/2(\partial Q/\partial\alpha_k)=0$，則得

$$\sum_j P_{kj}(\widetilde{Y}_{kj}-\alpha_k-\alpha_j)+\sum_i P_{ik}(\widetilde{Y}_{ik}-\alpha_i-\alpha_k)=0.$$

这里，$\widetilde{Y}_{ij}$ 是 Y_{ij} 与群体平均間的偏差。即

$$\widetilde{Y}_{ij}=Y_{ij}-\overline{Y}.$$

从而正規方程(normal equation)成为

$$2\alpha_k x_k+2\sum_j P_{kj}\alpha_j=2a_k x_k$$

$$(k=1, 2, \cdots, n). \tag{6.3}$$

其中，x_k 是 A_k 的频率：

$$x_k=\sum_j P_{kj},$$

而根据定义，a_k 值满足关系：

$$a_k=\sum_j P_{kj}\widetilde{Y}_{kj}/x_k. \tag{6.4}$$

而且象在最小二乘法的理論中所知道的，由含有 n 个参数 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 的回归直綫所减去的变差，就是正规方程的右边与参数推定值的乘积之和。所以把它叫做相加的遺傳变差 (additive genetic variance)，或基因变差 (genic variance)，如用 V_g 表示，則得下列的重要公式

$$V_g=2\sum_k \alpha_k a_k x_k. \tag{6.5}$$

現在回到(6.1)，d_{ij} 是表示：遺傳型值 Y_{ij} 与根据基因相加作用的期望值

$$\overline{Y}+\alpha_i+\alpha_j$$

間有多少差异，这是由于相同基因間的显隐性关系。在統計上，可以看作是相同基因間的相互作用。这种起因于显隐性关系的变差，叫做显性变差 (dominance variance)，如用 V_d 表示，則

$$V_d=\sum_{ij} P_{ij}d_{ij}^2.$$

在这里，V_d 就是最小二乘法上所謂殘差平方和，所以如把起因于遺傳型值的变差 (genotypic variance) 写作 V_y，則得

$$V_y=V_g+V_d.$$

习題 1 用正規方程(6.3)，証明 α_i，d_{ij} 等間有如下的关系：

$$\sum_i \alpha_i x_i=0, \quad \sum_i d_{ij}P_{ij}=\sum_j d_{ij}P_{ij}=0,$$

$$\sum_{ij}\alpha_i d_{ij}P_{ij}=0.$$

习題 2 設遺傳型 A_iA_j 的实际频率与完全随机交配下所期望的频率之比为 θ_{ij}，称之为交配的非随机系数(Kimura, 1957 a)，則

$$P_{ij}=\theta_{ij}x_ix_j,$$

对于 θ_{ij} 有下面的关系成立

$$\sum_i x_i\theta_{ij}=\sum_j x_j\theta_{ij}=1.$$

特别在 θ_{ij} 是定数时，

$$\Delta P_{ij}=\theta_{ij}\Delta x_i\cdot x_j+\theta_{ij}x_i\Delta x_j,$$

所以請証明可得

$$\begin{aligned}\Delta\bar{Y}&=\sum_{ij}\Delta P_{ij}\cdot Y_{ij}\\&=\sum_{ij}(\theta_{ij}x_j\Delta x_i+\theta_{ij}x_i\Delta x_j)(\bar{Y}+\alpha_i+\alpha_j+d_{ij})\\&=\sum_i\alpha_i\Delta x_i+\sum_j\alpha_j\Delta x_j=2\sum_k\alpha_k\Delta x_k.\end{aligned}$$

一般，α_i 可看作是群体平均的增加 $\Delta\bar{Y}$ 对 A_i 的頻率增加 Δx_i 的回归系数。仿照 Fisher 对二个等位基因的研究 (1930, 1941)，把 α_i 叫做 A_i 的平均效应 (average effect)，把(6.4)所标明的 a_i 叫做 A_i 的平均过剩(average excess)。再关于这个問題，可参照卷末的 Crow and Kimura (1956)。

习題 3　群体中的个体进行完全随机的交配时，$\theta_{ij}=1$，基因的平均效应与平均过剩相等，得

$$\alpha_i=a_i=\sum_j x_j\widetilde{Y}_{ij},$$

所以請証明

$$V_g=2\sum_i x_i\alpha_i^2=2\sum_i x_i a_i^2,$$

又

$$\sum_i d_{ij}x_i=\sum_j d_{ij}x_j=0.$$

习題 4　在植物育种学的实驗研究中，往往取二个純系 P_1 和 P_2，由此作成杂种第一代，以后每代由自家受精而继續繁殖。在这种場合，也可看作是二个等位基因 A 和 a，但两者的頻率始終是 0.5。在这种特殊情况下，采用 Mather (1949) 的下列記号，較为方便。設 AA，Aa 和 aa 的遺傳型值各为 d，h 和 $-d$，置

$$D=d^2,\quad H=h^2.$$

这时 F_2 的遺傳型变差 V_{F_2} 可表示为

$$V_{F_2}=\frac{1}{2}D+\frac{1}{4}H,$$

并驗証 $D/2=V_g$，$H/4=V_d$。有几个相互不連鎖的基因位点时，上述的 D 和 H 就分别作为各位点的 d^2 和 h^2 的值的和。Mather 把 D 称为可固定的

遺傳变差(fixable heritable variance),把 H 称为不能固定的遺傳变差(unfixable heritable variance)。

§7 近亲間的相关

把問題限于完全随机交配的群体，考慮一基因位点上有 n 个等位基因 $A_1, A_2, \cdots, A_n$ 的情况。在这群体中，取有特定近緣关系的二个体 I 和 J，設两者的遺傳型分别为 A_iA_j 和 $A_{i'}A_{j'}$。以群体平均为基准时，如前节所示，I 的遺傳型值 $\widetilde{Y}_{ij}$ 可表示为

$$\widetilde{Y}_{ij}=a_i+a_j+d_{ij}.$$

为什么不用 α_i，而用 a_i，这已在前节习題 3 中証明，因为这时两者是相等的。假使沒有因自然选擇而起的頻率变化，且外界环境保持恒定，則各遺傳型頻率及 a_i, d_{ij} 等的值，在世代之間沒有变化。

現設 I 与 J 間的遺傳型值間的协方差(covariance)为 W_{IJ}，則

$$\begin{aligned}W_{IJ}&=\mathrm{E}(\widetilde{Y}_{ij}\widetilde{Y}_{i'j'})\\&=\mathrm{E}\{(a_i+a_j+d_{ij})(a_{i'}+a_{j'}+d_{i'j'})\}\\&=\mathrm{E}(a_ia_{i'})+\mathrm{E}(a_ia_{j'})+\cdots+\mathrm{E}(a_id_{i'j'})+\mathrm{E}(a_jd_{i'j'})\\&\quad+\cdots+\mathrm{E}(d_{ij}d_{i'j'}).\end{aligned}$$

这里 E 表示在一切可能情况下所取的期望值。为便于說明起見，假定 I 个体从母体得到 A_i，从父体得到 A_j，同样，J 个体从母体得到 $A_{i'}$，从父体得到 $A_{j'}$。先来看

$$\mathrm{E}(a_ia_{i'})$$

这一项，因为

$$\mathrm{E}(a_i)=\mathrm{E}(a_{i'})=\sum_i a_ix_i=0,$$

所以当 A_i 和 $A_{i'}$ 来自完全独立的祖先时，这一項是 0。其次，若 A_i 和 $A_{i'}$ 来自祖先的同一基因（把它略記为 $A_i\equiv A_{i'}$），則 $a_i=a_{i'}$，这一項就等于

$$\sum_i x_i a_i^2=\frac{1}{2}V_g.$$

这里，V_g 是基因变差(参照前节习题3)。所以如 I 从母体得到的基因与 J 从母体得到的基因来自同一祖先基因的概率是 ϕ_{11}，则

$$E(a_i a_{i'}) = (\phi_{11}/2)V_g.$$

同样，如 I 的母系基因与 J 的父系基因来自祖先的同一基因的概率是 ϕ_{12}，I 的父系基因与 J 的母系基因，I 的父系基因与 J 的父系基因間的同样的概率分别是 ϕ_{21}，ϕ_{22}，則得

$$E(a_i a_{j'}) = (\phi_{12}/2)V_g,\quad E(a_j a_{i'}) = (\phi_{21}/2)V_g,$$
$$E(a_j a_{j'}) = (\phi_{22}/2)V_g$$

等。其次来看

$$E(a_i d_{i'j'})$$

这一項，它等于0. 关于这一点，象下面这样来思考，就容易了解。个体 J 内的二基因 $A_{i'}$ 和 $A_{j'}$ 在随机交配的关系上，是相互独立的，所以 $A_{i'}$ 和 $A_{j'}$ 都不可能与 A_i 来自祖先的共同基因。因而現在假定 A_i 和 $A_{j'}$ 是独立的，則得

$$E(a_i d_{i'j'}) = E(a_i \sum_{j'} d_{i'j'} x_{j'}),$$

注意到

$$\sum_{j'} d_{i'j'} x_{j'} = 0,$$

則这一項是0. 同样

$$E(a_j d_{i'j'}) = E(d_{ij} a_{i'}) = E(d_{ij} a_{j'}) = 0.$$

最后

$$E(d_{ij} d_{i'j'})$$

这一項，只有在 $A_i \equiv A_{i'}$，同时 $A_j \equiv A_{j'}$；或 $A_i \equiv A_{j'}$，同时 $A_j \equiv A_{i'}$ 的时候，才不是0，所以这一項等于

$$(\phi_{11}\phi_{22} + \phi_{12}\phi_{21})V_d.$$

其中 V_d 是显性变差，

$$V_d = \sum_{ij} P_{ij} d_{ij}^2 = \sum_{ij} x_i x_j d_{ij}^2.$$

总括上面的結果，得到下面的重要公式：

$$W_{IJ}=\frac{\phi_{11}+\phi_{12}+\phi_{21}+\phi_{22}}{2}V_g+(\phi_{11}\phi_{22}+\phi_{12}\phi_{21})V_d. \quad (7.1)$$

所以有一定亲緣关系的二个体 I 与 J 間的相关系数可由

$$r_{IJ}=W_{IJ}/V_T$$

求得。这里，V_T 是关于个体的全变差，設起因于环境的变差为 V_E，則

$$V_T=V_g+V_d+V_E.$$

探討多数基因位点时，必須考虑到位于各别位点上的基因間的相互作用，即上位現象(epistasis)，那就需要把(7.1)推广。

习題 1 当 I，J 有着第 5 节图 5.2 所示的兄弟关系时，

$$\phi_{11}=\phi_{22}=\frac{1}{2},\ \phi_{12}=\phi_{21}=0,$$

从而証明

$$W_{IJ}=\frac{1}{2}V_g+\frac{1}{4}V_d.$$

又对同一节图 5.3，5.4，5.5 所示的父子，半兄弟，表兄弟的关系，求 W_{IJ}。

习題 2 等位基因仅有二个时，設它們是 A 和 a，群体中两者的频率是 p 和 $q(p+q=1)$，則三个遺傳型 AA，Aa 和 aa 的频率各为 p^2，$2pq$ 和 q^2。又設这些遺傳型值为 i，j 和 k，取群体平均为 $0(ip^2+2jpq+kq^2=0)$，証明基因变差是

$$\begin{aligned}V_g&=2pq\{(i-j)p+(j-k)q\}^2\\&=(p^2i-q^2k)^2/(2pq),\end{aligned}$$

显性变差是

$$V_d=p^2q^2(i-2j+k)^2.$$

其次，作成亲子、兄弟間的遺傳型値間的相关表，由此直接算出协方差，証明习題 1 所得的結果。

[提示：对于亲子的关系，

$$W_{IJ}=p^3i^2+2p^2qij+pq(p+q)j^2+2pq^2jk+q^3k^2.]$$

第2章　自然选擇的作用

群体中許多基因在分离时，通过有性生殖，这些基因間发生結合、再結合，每代产生很多遺傳型。自然群体中遺傳型数目究竟有多大，象下面这样去推想，就容易想象。現在暫且假定含有二种等位基因的位点有 1000 个，則在完全随机交配下，可能出現的遺傳型种类是 3^{1000} 种，这大致上等于 10^{477}. 而可観宇宙中存在的电子总数，据估計在 10^{100} 以下，可見遺傳型种类确实是天文学上的数字。各遺傳型都有各自特有的生存率、增殖力，所以群体的遺傳构成随着发生变化。还有，环繞着这群体的物理的和生物的环境是在不断地変化、变动着的，所以实际上，在群体里所发生的遺傳組成的变化中，即使单是起因于自然选擇的，一定也是非常复杂的。但是不把这过程解释明白，要了解生物进化的机制，那是不能想望的。本章中，假定群体相当大，展开关于自然选擇的决定論的探討。

§6　世代构造的数学模型

1. 不連續模型　象在一年生植物的群体中所見到的，群体中的个体都属于同一世代，在下一次繁殖期到来以前，属于上一世代的个体完全为新世代的个体所代替。在这种情况，表示时間的参数 t 取整数値 0, 1, 2, …. 在这个模型里，各遺傳型的适应度 (fitness) 用属于这遺傳型的任一个体所留的子女的平均数来表示。計算这平均数时，要除去发育中途死亡的个体，只統計活到繁殖期的孩子数，这样最为适当。以下用符号 w 表示适应度，依据 Wright，称之为选擇值(selective value)(例如参照 Wright,1955)。

选择值的群体平均数 $(\bar{w})$ 比 1 大时，群体每代扩大，反之，比 1 小时，每代縮小。$\bar{w}=1$ 的群体，在数目上可以說是处于定常状态。設遺傳型 A_iA_j 的选擇值为 w_{ij}，探討一下这个值与 1 之差是有好处的，称这个差 $s_{ij}=w_{ij}-1$ 为 A_iA_j 的选擇系数 (selection coefficient)。这时通常以某一特定遺傳型为基准，便宜上取它的选擇值为 1. 现在来考虑含有任何数目的等位基因的位点，設它的第 i 个基因 A_i 的相对頻率为 x_i。如在下一个世代时，这个頻率为 x_i'，則根据上面的定义，得

$$x_i'=\sum_j P_{ij}w_{ij}/\sum_{ij} P_{ij}w_{ij},$$

所以，設每一代 x_i 的变化为 Δx_i，則得

$$\Delta x_i=x_i'-x_i=x_i(w_{i.}-\bar{w})/\bar{w}. \tag{8.1}$$

其中

$$w_{i.}=\sum_j w_{ij}P_{ij}/x_i, \quad \bar{w}=\sum_{ij} w_{ij}P_{ij}=\sum_i w_{i.}x_i.$$

P_{ij} 是上一章說明的順序性遺傳型 A_iA_j 的相对頻率。

2. 連續模型 象在人类群体中所看到的，在群体中屬于各別世代的个体往往共存，现在就来考虑这样的例子。这里，必須把表示世代的参数 t 当作連續型变数来处理，严密的数学处理相当麻煩，但在总的方面，重要的是到达一定年龄分布状态的各遺傳型的几何級数增减率，可定义如下。設任意的 A_iA_j 个体活到年龄 x 的概率为 $l_{ij}(x)$，又在这个年龄时的出生率是 $b_{ij}(x)$，适应度 m_{ij} 是方程

$$\int_0^{\infty} e^{-m_{ij}x}\, l_{ij}(x)\, b_{ij}(x)\, dx=1 \tag{8.2}$$

的实根。$l_{ij}(x)\,b_{ij}(x)$ 决不为負的，故上式左边的积分在 m_{ij} 为实数时是 m_{ij} 的单調函数，可取 0 到 $+\infty$ 間的任何值。所以滿足 (8.2) 的 m_{ij} 的实根（暂且假設为 a_{ij}）有一个，而且只有一个。在此以外 (8.2) 有无数个复根，它們的实数部分常比 a_{ij} 小，所以群体应当稳

定在一定的年龄分布上。根据 Fisher，称 a_{ij} 为 Malthus 参数 (Malthusian parameter)。雌雄間繁殖力不同时，可取两者的平均。a_{ij} 的群体平均 $\bar{a}$ 为正时，群体趋向增大，而 $\bar{a}$ 为負时，則在縮小。$\bar{a}$ 等于 0 时，数目上处于平衡状态。现在設群体中順序性遺傳型 A_iA_j 的个体数为 n_{ij}，則用 a_{ij} 表示它的几何級数的增減率，所以可置

$$\frac{1}{n_{ij}}\frac{dn_{ij}}{dt}=a_{ij}.$$

为計算方便計，置

$$N_i=\sum_j n_{ij},\quad N=\sum_{ij} n_{ij}=\sum_i N_i,$$

則

$$\frac{dx_i}{dt}=\frac{d}{dt}\left(\frac{N_i}{N}\right)=\left(\frac{dN_i}{dt}N-\frac{dN}{dt}N_i\right)\Big/N^2$$

$$=\frac{N_i}{N}\frac{1}{N_i}\frac{dN_i}{dt}-\frac{N_i}{N}\frac{dN}{dt}\frac{1}{N}$$

$$=x_i\left(\frac{1}{N_i}\frac{dN_i}{dt}-\frac{1}{N}\frac{dN}{dt}\right).$$

但在这里若注意到

$$\frac{dN_i}{dt}=\frac{d}{dt}\sum_j n_{ij}=\sum_j a_{ij}n_{ij}$$

$$\frac{dN}{dt}=\frac{d}{dt}\sum_{ij} n_{ij}=\sum_{ij} a_{ij}n_{ij}$$

可得 x_i 的时間变化率公式

$$\frac{dx_i}{dt}=x_i(a_{i\cdot}-\bar{a}).\qquad(8.3)$$

这里

$$a_{i\cdot}=\sum_j a_{ij}P_{ij}/x_i,\quad \bar{a}=\sum_{ij} a_{ij}P_{ij}=\sum_i a_{i\cdot}x_i$$

$(n_{ij}=NP_{ij})$。把 (8.3) 的右边所出現的量 $a_{i\cdot}-\bar{a}$ 写作 $\tilde{a}_{i\cdot}$，称它为基因 A_i 对适应度的平均过剩 (average excess)。至于有二种等位基因时 ($i=1, 2$)，Fisher (1930, 1941) 以 a 代替 $a_{1\cdot}-a_{2\cdot}$ 的值，把它叫做用 A_1 置換 A_2 时的平均过剩 (参照 Crow and Kimura,

1956)。(8.3)也可写作

$$\frac{d}{dt}\log x_i=\tilde{a}_i, \quad (8.4)$$

由此得下列法則，即基因 A_i 的頻率的对数增加率等于那个基因的适应性的平均过剩。

3. 連續与不連續两模型間的关系 上述两个理論的模型都是从現实得来的抽象，各有长处和缺点。如上所述，不連續模型适于具有一年生植物那样的不連續世代結构的生物群体的处理，一般得到(8.1)式的差分方程。这不是綫性的，而且分母 $\bar{w}$ 也不是普通常数，所以不能希望有明显的解析形式的答案，但对于特定場合的数值計算，却有不少的便利。一方面，連續模型适于人类群体那样，生殖和死亡在一年四季連續发生的場合的处理，得到(8.3)那样的一阶微分方程。其中沒有相当于上面(8.1)的分母 $\bar{w}$ 的值，还可自由地运用微积分学，所以在解析处理上一般比不連續模型容易。自然群体的进化一般极为緩慢，由不連續模型得到的差分方程也可近似地用微分方程代替，又即使在不連續模型中，如采用一个世代的平均长度作为时間单位，可得便利。概括而言，同一問題由以上二个模型处理时，結論很少有本质上的差异。不过这里应当注意，两模型中适应度的尺度初看是不同的。Malthus 参数 a 大致等于选擇值的对数 $\log w$。由不連續模型的方程(8.1)向連續模型的方程(8.3)的变換，可按下列形式进行。設

$$w_{ij}=1+s_{ij}\Delta t,$$

代入(8.1)，則得

$$\frac{\Delta x_i}{\Delta t}=\frac{x_i(s_{i\cdot}-\bar{s})}{1+\bar{s}\Delta t},$$

所以这里如 $\Delta t\to 0$，則得与(8.3)同样的方程。这方程就是用选擇系数 s_{ij} 代替了 Malthus 参数 a_{ij} 的結果，因此可以知道，在这个意义上，連續模型的 Malthus 参数与不連續模型的选擇系数間，存

在着饒有趣味的对应关系。

习題1 用不連續模型,考慮一对等位基因 A 和 A',設两者的频率各为 p 和 q。如交配完全随机,三遺傳型 AA, AA' 和 $A'A'$ 的选擇值为 1, $1-s$, $1-2s$, 証明

$$\Delta p = spq/(1-2sq) \quad (q=1-p).$$

又这公式用微分方程

$$\frac{dp}{dt}=\frac{spq}{1-2sq}$$

代替,则以 s 为常数,可得

$$st=(1-2s)\log p-\log q+C.$$

于是設初始频率为 p_0,試确定常数 C,并就 $s=0.01$, $p_0=0.01$ 时,用图表示基因頻率的时間变化。其次,用連續模型,設三遺傳型的适应度分别为 0, $-\alpha$, -2α, 証明

$$\frac{dp}{dt}=\alpha pq,$$

或

$$\frac{d}{dt}\log\left(\frac{p}{q}\right)=\alpha.$$

习題2 在上一問題中,假定基因 A 与 A' 間在适应度上沒有显隐性关系。这里,来探討一下 A 对 A' 有显性的例子。用不連續模型,設 AA 和 AA' 的选擇值为 1, $A'A'$ 的选擇值为 $1-s$, 则当交配完全随机时,可得

$$\Delta p=\frac{spq^2}{1-sq^2}.$$

从而,証明在隐性基因的频率低时 $(q\ll 1)$,选擇的效果显著减弱。用

$$\frac{dp}{dt}=\frac{spq^2}{1-sq^2}$$

代替上式,可得

$$st=(1-s)\log p-\log q+(1/q)+C,$$

这里,C 可由基因的初始频率来决定。其次,用連續模型,設 AA 与 AA' 的适应度为 0, $A'A'$ 的适应度为 $-\alpha$, 则得

$$\frac{dp}{dt}=\alpha pq^2.$$

习題3 表示隐性性状的个体每代完全除去时,则在問題2中,$s=1$,所以得

$$\Delta p=\frac{q^2}{1+q}.$$

这时，設第 t 代的 A' 的频率为 q_t，则

$$1/q_t=1+1/q_{t-1},$$

从而証明

$$q_t=1/(t+1/q_0) \quad (t=0, 1, 2, \cdots).$$

用这公式，如由隐性基因而起的不良性状所有者，其比例为每一万人中有一个，試問通过选擇作用，使这基因频率下降一半需要多少年数。設一世代的长为 30 年。

习題 4 等位基因 A 与 A' 間，不論有那一种显隐性关系，如三遺傳型 AA, AA' 和 $A'A'$ 的选擇值为 $1, 1-hs, 1-s$，则在完全随机的交配下，証明

$$\Delta p=\frac{spq[h(p-q)+q]}{1-2hspq-sq^2},$$

試用这公式証明 Wright 的公式

$$\Delta p=\frac{pq}{2}\frac{1}{\bar{w}}\frac{d\bar{w}}{dp}.$$

这里，$\bar{w}$ 是选擇值的群体平均。

§9 群体适应度随自然选擇的增加

群体中存在有遺傳的变异时，由于自然选擇的作用，适应性高的应該增加，所以直覺的想来，群体的平均适应度也应增加，而且它的增加率是跟遺傳的变异量成比例的。但是把这个問題严密地立出数学公式来，不一定容易，自著名的 Fisher(1930) 的自然选擇基本定理(fundamental theorem of natural selection) 发表以来，直到今日，各种議論不断出現 (Wright 1955, Crow 1955, Crow and Kimura 1956)。幸而作者对这問題得到了稍可滿意的結果。以下談一談这方面的情况，就把既得結果作为一个特例。先就一个基因位点来看。設有任意的 n 个等位基因 A_1, A_2, $\cdots$, A_n，記 A_i 的相对頻率为 x_i。如順序性遺傳型 A_iA_j 的頻率为 P_{ij}，則适应度 a_{ij} 可用下式表示：

$$a_{ij}=\bar{a}+\alpha_i+\alpha_j+d_{ij}.$$

这里，$\bar{a}$ 是适应度的群体平均，α_i 和 α_j 分别为 A_i 和 A_j 对适应度

的平均效应，而 d_{ij} 是由显隐性关系引起的偏差。把适应度看做一个数量性状，则在第1章第6节中，用 a_{ij} 代替 Y_{ij}，从而就可决定这些数值。就是使

$$Q=\sum_{ij} P_{ij}(\tilde{a}_{ij}-\alpha_i-\alpha_j)^2$$

的值最小，以求得平均效应和显性偏差 (dominance deviation)。这里 $\tilde{a}_{ij}$ 表示 $a_{ij}-\bar{a}$。以下为了要表示与平均数的偏差，一般都附加波状符号（～）。从而与(6.3)及(6.5)相对应，得到

$$2\alpha_k x_k+2\sum_j P_{kj}\alpha_j=2x_k\tilde{a}_{k\cdot} \quad (k=1, 2, \cdots, n) \tag{9.1}$$

及

$$V_g=2\sum_k \alpha_k\tilde{a}_{k\cdot}x_k. \tag{9.2}$$

这里，V_g 是对适应度的基因变差。根据(8.3)，可将(9.2)写作

$$V_g=2\sum_k \alpha_k \frac{dx_k}{dt}. \tag{9.3}$$

如第6节习题3所示，在完全随机交配下，平均效应 α_i 等于平均过剩 $\tilde{a}_{i\cdot}$，在这場合，特别有

$$V_g=2\sum_i x_i\tilde{a}_{i\cdot}^2. \tag{9.4}$$

但是如导入同一节习题2所讲的交配中的非随机系数

$$\theta_{ij}=P_{ij}/x_ix_j \quad (\theta_{ij}=\theta_{ji}), \tag{9.5}$$

则从(9.1)得

$$\alpha_k+\sum_j x_j\theta_{kj}\alpha_j=\tilde{a}_{k\cdot}$$

或

$$\sum_j x_j\theta_{ij}\tilde{a}_{ij}=\alpha_i+\sum_j \alpha_j\theta_{ij}x_j. \tag{9.6}$$

群体的平均适应度是

$$\bar{a}=\sum_{ij} P_{ij}a_{ij},$$

所以考虑它的时間变化，则得

$$\frac{d\bar{a}}{dt}=\sum_{ij}\frac{da_{ij}}{dt}P_{ij}+\sum_{ij}a_{ij}\frac{dP_{ij}}{dt}.$$

或者象物理学上常用的那样，在文字上加点，用来表示对时間的微商，則得

$$\dot{\bar{a}}=\sum_{ij}\dot{a}_{ij}P_{ij}+\sum_{ij}a_{ij}\dot{P}_{ij}. \tag{9.7}$$

右边第一項是各遺傳型适应度的时間变化率的群体平均，把这一項表示为

$$\frac{\overline{da}}{dt}$$

或再簡单地表示为

$$\bar{\dot{a}}.$$

其次，右边第二項为

$$\sum_{ij}a_{ij}\dot{P}_{ij}=\sum_{ij}(\bar{a}+\tilde{a}_{ij})\dot{P}_{ij}=\bar{a}\sum_{ij}\dot{P}_{ij}+\sum_{ij}\tilde{a}_{ij}\dot{P}_{ij},$$

这里

$$\sum_{ij}P_{ij}=1,$$

从而如注意到

$$\sum_{ij}\dot{P}_{ij}=0,$$

則这項成为

$$\sum_{ij}\tilde{a}_{ij}\dot{P}_{ij}.$$

再，代入 $P_{ij}=x_ix_j\theta_{ij}$，則这一項成为

$$\sum_{ij}\tilde{a}_{ij}\dot{x}_ix_j\theta_{ij}+\sum_{ij}\tilde{a}_{ij}x_i\dot{x}_j\theta_{ij}+\sum_{ij}\tilde{a}_{ij}x_ix_j\dot{\theta}_{ij}. \tag{9.8}$$

把(9.6)的关系代入上式第一項，則得

$$\sum_i\dot{x}_i(\alpha_i+\sum_j\alpha_j\theta_{ij}x_j)=\sum_i\alpha_i\dot{x}_i+\sum_{ij}\alpha_j\dot{x}_ix_j\theta_{ij},$$

但从(9.3)可以知道，这式的右边第一項是 $V_g/2$，又从定义知

$$\sum_i P_{ij}=x_j,\ 或\ \sum_i x_ix_j\theta_{ij}=x_j,$$

所以

$$\sum_i x_i\theta_{ij}=1,$$

把这式的两边对时間微分,得到

$$\sum_i \dot{x}_i\theta_{ij}+\sum_i x_i\dot{\theta}_{ij}=0.$$

从而右边第二項成为

$$\sum_{ij}\alpha_j\dot{x}_i x_j\theta_{ij}=\sum_j \alpha_j x_j\left(-\sum_i x_i\dot{\theta}_{ij}\right)$$
$$=-\sum_{ij}\alpha_j x_i x_j\dot{\theta}_{ij}.$$

同样,(9.8)的右边第二項等于

$$\frac{1}{2}V_g-\sum_{ij}\alpha_i x_j x_i\dot{\theta}_{ij}.$$

所以(9.8)全体可写作

$$V_g+\sum_{ij}(\tilde{a}_{ij}-\alpha_i-\alpha_j)x_i x_j\dot{\theta}_{ij},$$

或

$$V_g+\sum_{ij}d_{ij}P_{ij}(\dot{\theta}_{ij}/\theta_{ij}).$$

总結以上結果,得到

$$\frac{d\bar{a}}{dt}=V_g+\overline{\frac{da}{dt}}+\sum_{ij}P_{ij}d_{ij}\frac{d}{dt}\log\theta_{ij}. \tag{9.9}$$

从而,群体的适应度的增加率可分解为上式右边所示的三个成分。即(1)适应度的基因变差:由(9.3),这是

$$V_g=2\sum_k\alpha_k\frac{dx_k}{dt},$$

略去(9.9)右边第二項以下的部分,得到

$$\dot{\bar{a}}=2\sum_k\alpha_k\dot{x}_k,$$

这可看作是适应度增加率对基因頻率增加率的回归式,所以这一項可以看作是种的基因构成改善后所引起的适应度的增加。在这个意义上,第一項在遺傳进化学上是最为重要的,而 Fisher 的所謂自然选擇基本定理

$$\frac{d\bar{a}}{dt}=V_g, \tag{9.10}$$

可以說是生物进化的基本性质的抽象化。(2)各遺傳型的适应度变化率的群体平均：这一項可以看作是表示因物理的或生物的环境变化而引起的群体适应度的变化。各遺傳型的适应度是常数时，这一項是 0. (3)交配的非随机系数的对数变化率与显性偏差的乘积的群体平均：这一項是由于交配系与等位基因相互作用(allelic interaction)之間的共同作用而引起的适应度增加，可以看作是起因于遺傳的环境变化部分。当交配的进行完全随机时，对所有的 i 和 j 値，都有 $\theta_{ij}=1$，所以这一項等于 0. 又当等位基因間沒有显隐性关系时，$d_{ij}=0$，所以这时第三項也等于 0.

以上单是考虑一个基因位点，但含有任何数目的基因位点时，一般可把(9.9)推广，由此可把基因間的相互作用，特别是上位現象在进化上的意义，列成公式，这点是很重要的。有兴趣的讀者請参閱卷末文献所列的作者論文(Kimura, 1957a)。

§10 自然选擇的基本定理及其边界

在上节所推导的一般式(9.9)中，单是考虑右边第一項，則成

$$\frac{d\bar{a}}{dt}=V_g, \tag{10.1}$$

这是 Fisher 的所謂自然选擇的基本定理。右边所表現的 V_g 是群体适应度中基因(或相加)的变差，所以决不会是負的，这就意味着

$$\frac{d\bar{a}}{dt}\geqslant 0. \tag{10.2}$$

假使认为(9.9)的右边第二項和第三項是表示由广义的环境(包括物理的、生物的、遺傳的环境)所引起的效应，那末可以解釋为，在固定的外界环境中，群体适应度常常随自然选擇而趋向增加。Fisher (1930)认为这是生物学上的基本法則，可与热力学的第二定律相比。但是在現实的生物群体中，由于外界物理条件的变化、不同种或不同遺傳型間的竞爭等原因，(9.9)的右边第二項和第三

項一般将不等于0。下面考察几个富有趣味的例子。

(1) 各遺傳型的适应度是不依賴于时間或基因頻率的常数时，(9.9)的右边第二項成为0，得

$$\frac{d\bar{a}}{dt}=V_g+\sum_{ij}P_{ij}d_{ij}\frac{d}{dt}\log\theta_{ij}. \tag{10.3}$$

現在設有常数 λ_{ij}，对所有的 i，j，滿足下面的遺傳型頻率之間的关系：

$$P_{ij}^2=\lambda_{ij}P_{ii}P_{jj}, \tag{10.4}$$

則(10.3)的右边第二項成为0，得到与 Fisher 的自然选擇基本定理的同一形式。这是因为，把 $P_{ij}=x_ix_j\theta_{ij}$ 代入(10.4)，得到

$$\theta_{ij}^2=\lambda_{ij}\theta_{ii}\theta_{jj},$$

从而有

$$2\frac{d}{dt}\log\theta_{ij}=\frac{d}{dt}\log\theta_{ii}+\frac{d}{dt}\log\theta_{jj},$$

如把它代入(10.3)的右边第二項，則得

$$\sum_{ij}P_{ij}d_{ij}\frac{d}{dt}\log\theta_{ij}=\frac{1}{2}\sum_{ij}d_{ij}\left(\frac{d}{dt}\log\theta_{ii}+\frac{d}{dt}\log\theta_{jj}\right)P_{ij}$$

$$=\frac{1}{2}\sum_i\frac{d}{dt}\log\theta_{ii}\sum_j d_{ij}P_{ij}+\frac{1}{2}\sum_j\frac{d}{dt}\log\theta_{jj}\sum_i d_{ij}P_{ij},$$

但

$$\sum_j d_{ij}P_{ij}=\sum_i d_{ij}P_{ij}=0,$$

所以(§6，习題1)容易了解，(10.3)的右边第二項是0。Fisher 对二个等位基因的場合，用完全不同的方法得到了同样的証明。

(2) 象在物理学上所周知的一样，在保守力場中存在有势函数。与这函数相应，Wright (1955) 推究了內在选擇值 (internal selective value) 的函数。作者用独立的方法，导入了內在适应度的函数。考虑基因頻率的微小变化的一次式

$$2\sum_{i=1}^{n}\alpha_i dx_i \quad \left(\sum_{i=1}^{n}x_i=1\right),$$

当上式恰等于某函数 I 关于基因頻率 $(x_1, x_2, \cdots, x_n)$ 的全微分时，即称函数 I 为某基因系的內在适应度 (internal fitness)。如果这样的函数 I 存在，則

$$dI=2\sum_i \alpha_i dx_i,$$

所以

$$\frac{dI}{dt}=2\sum_i \alpha_i \frac{dx_i}{dt}=V_g,$$

即群体的內在适应度的增加率等于那时的基因变差。若內在适应度与群体的平均适应度之差是常数，則

$$\frac{dI}{dt}=\frac{d\bar{a}}{dt},$$

所以在这个場合，也得到 Fisher 基本定理的同一形式。但是反过来不一定正确，內在适应度的存在不是 Fisher 基本定理成立的必要条件。

(3) 遺傳型間有竞爭时，屡次出現奇妙的現象。举其一例，設一基因位点有三个复等位基因 A_1, A_2 和 A_3，現在来看一看象下面这样的情况。令这三个基因的相对頻率各为 x_1, x_2 和 x_3，为簡单起見，假定沒有显隐性关系。或者象在蘚类中所見的一样，单相世代发达，复相世代只有寄生的份儿，探討这样的生物群体是比較方便的。現在如 A_1 个体为 A_2 个体所包圍，适应度減少 c，反之，如由 A_3 包圍，增加 c。同样，如 A_2 由 A_3 包圍，适应度減少 c，由 A_1 包圍，增加 c。最后，如 A_3 由 A_1 包圍，适应度減少 c，由 A_2 包圍，增加 c (参照图 10.1)。各遺傳型为同种个体包圍时，如它們的适应度用 Malthus 系数来測量，則等于 0，因此表示基因頻率变化的方程为

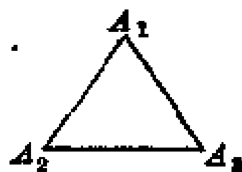


图 10.1

$$\left.\begin{aligned}\frac{dx_1}{dt}&=cx_1(x_3-x_2),\\ \frac{dx_2}{dt}&=cx_2(x_1-x_3),\\ \frac{dx_3}{dt}&=cx_3(x_2-x_1).\end{aligned}\right\}\qquad(10.5)$$

如以坐标系中的一点来表示群体的基因频率，则可証明，显示变化历程的軌道是一条閉曲綫。由(10.5)，得

$$\begin{aligned}\frac{d}{dt}\log(x_1x_2x_3)&=\frac{1}{x_1}\frac{dx_1}{dt}+\frac{1}{x_2}\frac{dx_2}{dt}+\frac{1}{x_3}\frac{dx_3}{dt}\\&=c(x_3-x_2)+c(x_1-x_3)+c(x_2-x_1)=0,\end{aligned}$$

从而

$$x_1x_2x_3=C_0.\qquad(10.6)$$

这里 C_0 是由初始条件所决定的常数。又有

$$x_1+x_2+x_3=1.\qquad(10.7)$$

所以軌道是平面(10.7)与曲面(10.6)相交的閉曲綫。回轉的方向随 c 的正負，或向左旋，或向右旋，表示群体中基因組成的点却永远在这閉曲綫上継續运动。在这个場合，各基因的平均效应是

$$\begin{aligned}\alpha_1&=\tilde{a}_{1.}=c(x_3-x_2),\\ \alpha_2&=\tilde{a}_{2.}=c(x_1-x_3),\\ \alpha_3&=\tilde{a}_{3.}=c(x_2-x_1),\end{aligned}$$

群体的平均适应度始終是0，沒有变化。所以

$$\frac{d\bar{a}}{dt}=0.$$

另一方面，群体的基因变差是

$$V_g=c^2\{x_1(x_3-x_2)^2+x_2(x_1-x_3)^2+x_3(x_2-x_1)^2\},$$

除 $x_1=x_2=x_3=1/3$ 的特殊場合以外，一般不等于0。这个事实看来好象和(10.1)所示的 Fisher 基本定理相矛盾，但是实际上，看一看一般的(9.9)，它的右边第二項常常和第一項对消，

$$\frac{\overline{da}}{dt}=-V_g,$$

又第三項是0，所以全体应該为

$$\frac{d\bar{a}}{dt}=V_g+\frac{\overline{da}}{dt}+0=0,$$

与前一致。

习題1 設在一个基因位点上，有 n 个等位基因 $A_1, A_2, \cdots, A_n$，考察各遺傳型的适应度是常数的例子。若交配完全随机，試直接証明，群体的平均适应度决不因自然选擇而減少。[提示：設 A_iA_j 的适应度为 a_{ij}，則

$$\bar{a}=\sum_{ij}a_{ij}x_ix_j \quad (a_{ij}=a_{ji}).$$

故

$$\frac{d\bar{a}}{dt}=2\sum_{ij}a_{ij}\frac{dx_i}{dt}x_j=2\sum_{ij}a_{ij}x_i(a_{i.}-\bar{a})x_j$$

$$=2\sum_i x_i(a_{i.}-\bar{a})^2\geqslant 0.]$$

习題2 在上題同样条件下，用不連續模型，設 A_iA_j 的选擇植是 w_{ij}，試証明

$$\Delta\bar{w}=V_g/\bar{w}.$$

习題3 設一对基因 A 和 A' 的频率各为 x 和 $1-x$，三个遺傳型 AA, AA' 和 $A'A'$ 的适应度各为 $2e^{-kx}$, e^{-kx} 和 $0\,(k>0)$，問群体的平均适应度如何变化?

习題4 进行內交配时，設群体的內交系数为 F，則交配的非随机系数可表示为

$$\theta_{ii}=F/x_i+1-F, \quad \theta_{ij}=1-F \quad (i\neq j),$$

从而証明(9.9)的右边第三項可以象

$$\sum_{ij}x_ix_jd_{ij}\frac{d\theta_{ij}}{dt}=-\frac{(dF/dt)}{1-F}\sum_i d_{ii}x_i-F\sum_i d_{ii}\frac{dx_i}{dt}$$

一样，分解为由內交系数的变化所引起的部分以及起因于基因频率的变化的部分。

§11 起因于自然选擇的基因頻率的平衡及其稳定性

先来研究一个基因位点，設二个等位基因 A 和 A' 的相对頻

率各为 x 和 $1-x$。交配完全随机时，三种遺傳型 AA，AA' 和 $A'A'$ 的相对頻率分别为 x^2，$2x(1-x)$ 和 $(1-x)^2$，設这三种遺傳型的适应度各为 a，b 和 c，則由 (8.3)，x 的变化率可写作

$$\frac{dx}{dt}=\alpha x(1-x). \tag{11.1}$$

这里

$$\alpha=(a-b)x+(b-c)(1-x).$$

基因頻率不起变化之点，即

$$\frac{dx}{dt}=0 \tag{11.2}$$

的 x 值，叫做基因 A 的頻率的平衡点。从 (11.1) 可以知道，$x=0$ 及 $x=1$ 这两点是平衡点，但是如果注意到在没有变化之点自然选擇不能发生作用，就知这二点当然是平衡点，所以通常把它們看做“平凡平衡点”(trivial equilibrium)，而不予考虑。其次，来看一看 (11.1) 右边所出現的 α。如 α 在 x 的全范圍內始終是正的或始終是負的，那末只要 $0<x<1$，总是

$$\frac{dx}{dt}>0 \text{ 或 } \frac{dx}{dt}<0,$$

从而得到

$$\lim_{t\to\infty} x=1 \text{ 或 } \lim_{t\to\infty} x=0.$$

換言之，在这样的場合，只存在平凡平衡点，所以一方的基因最后完全把它方的基因排除。$a\geqslant b>c$，$a>b\geqslant c$，$a\leqslant b<c$，$a<b\leqslant c$ 等場合都属于这种情形。其次来看一看可以使 $\alpha=0$ 的例子。設 a，b，c 是常数，因为

$$\alpha=b-c+(a-2b+c)x, \tag{11.3}$$

所以如假定 $a-2b+c\neq0$，則滿足 $\alpha=0$ 的 x 值仅有一个，設它是 $\hat{x}$，則

$$\hat{x}=\frac{b-c}{(b-a)+(b-c)}. \tag{11.4}$$

在这場合，若 $b=c$，則 $\hat{x}=0$，若 $b=a$，則 $\hat{x}=1$，仍然成为“平凡”

平衡。我們有兴趣的是 $a<b>c$ 和 $a>b<c$ 这两个場合。現在設跟平衡点的偏差为 ξ，则

$$x=\hat{x}+\xi,$$

从而(11.1)成为

$$\frac{d\xi}{dt}=x(1-x)(a-2b+c)\xi. \qquad (11.5)$$

先来看一看 $a<b>c$，即异型接合体的适应度比任何一个同型接合体都高的例子。这种情况在遺傳学上叫做超显性(over-dominance)。这时

$$x(1-x)(a-2b+c)<0,$$

所以从(11.5)可以了解，当 ξ 是正的时候，ξ 趋向减少，反之，当 ξ 是負的时候，则趋向增加。因此，不論从什么值出发，只要不是0或1，A 的頻率一代代地接近于平衡頻率 $\hat{x}$。因为已經达到平衡点的頻率对离开平衡点的扰动有不断地回到原来位置的趋向，所以这个平衡是稳定的(stable)。应用 Lerner(1954)的术語，那就是群体在这个状态中能进行遺傳的自体調节(genetic homeostasis)。其次，来看一看 $a>b<c$，即异型接合体的适应度比任何一个同型接合体都低的例子。这时

$$x(1-x)(a-2b+c)>0,$$

所以与以前相反，ξ 是正时趋向增加，ξ 是負时趋向减少。从而，若 $x>\hat{x}$，最后有 $x=1$，若 $x<\hat{x}$，最后有 $x=0$。所以 $x=\hat{x}$ 的平衡即使受到微細的外界扰动立即会遭到破坏，平衡是不稳定的(unstable)。最后，$a=b=c$ 时，三遺傳型的适应度間沒有差别，从而基因頻率不因自然选擇而起变化。換言之，对任何的 x 值，都有

$$\frac{dx}{dt}=0,$$

在这意义上，任何的点都是平衡点。这样的平衡称为中性的

(neutral)。

稳定的平衡有永存性，在这种状态的基因系，通过悠长的进化历程，在群体中集积，更有可能成为种的特性，所以在遺傳进化学上有重大的意义。关于遺傳的性决定机制，以及自家不亲和基因系等支配交配系的基因，都存在着饒有兴味的問題，但这里从略。最后，群体适应度因自然选擇而有变化，不妨据此来对基因系进行分类。第一是，自然选擇使基因頻率变化，随着群体适应度增加，所以从自然选擇是使适应度高的基因构成留存下来的一点来看，这在常識上是最容易理解的。作者暂且把这样的基因系叫做正規的(regular)。第二是，自然选擇使基因頻率变化，随着群体适应度减低，把这样的基因系叫做退化的(degenerative)。第三是，尽管同样的頻率变化，但群体的适应度沒有变化，把它叫做无效的(vain)基因系。无效的基因系在各遺傳型的适应度間完全沒有差别，所以应当注意，这和自然选擇不起作用的中性(neutral)基因系不同，基因頻率由于自然选擇的作用而是有变化的。

习題 1 当等位基因有二个，而交配完全随机时，試証明平衡为稳定或不稳定，可根据

$$\frac{d}{dx}\left(\frac{dx}{dt}\right) \text{ 或 } \frac{d}{dx}\left(\frac{d}{dt}\log\frac{x}{1-x}\right)$$

的平衡点的正負来决定，把这方法应用于(11.1)的場合。又如取

$$a=(a-b)x+(b-c)(1-x)$$

为纵軸，取 x 为横軸繪出图象，試証明，倘使曲綫把 x 軸由左上向右下横切，則平衡是稳定的，假使由左下向右上横切，則平衡是不稳定的。若 a, b, c 是常数，則图表示直綫。

习題 2 一位点上有任何数目的等位基因，各遺傳型的适应度是常数，交配完全随机，試討論平衡的稳定性。[提示：利用基因系是正規的。欲知細节，請参照卷末所列的作者論文(1956 c)。]

习題 3 本节中，从有关自然选擇的作用的观点，把基因系分为正規的、退化的和无效的三种。到現在为止所討論过的各种基因系中，选出属于各該类别的具体例子。

第3章 突变的效应

基因除了每代复制与自己同一結构的基因，即具备所謂自己增殖(self-reproduction)的机能以外，还以低的频率变化成作用不同的基因，即具有所謂突变(mutation)的性质。每个基因的突变率是不同的，在迄今所研究的例子中，突变率大多是在 $10^{-5}\sim10^{-6}$ 的范圍內。由突变所生的基因，每代正确地复制这变化了的性质，这样地增殖着，但它在群体中的繁殖率通常与突变前不同，特别是在野生型基因(wild type gene)的場合，由突变所产生的突变基因(mutant gene)对自然选择大多比原来基因不利。这从自然选择能使基因系进化的观点看来，当然是所期望的。因为可以想象，在各个基因位点上，在悠长时間內，有的突变屡次发生，这些突变都經过自然选择的篩选，使适应度高的基因成为野生型基因，确立于現在的野生群体中。已如上述，突变率一般极低，所謂突变压(mutation pressure)本身是不能支配进化的过程的。还不如說，在起因于自然选择的进化中，对于提供进化的素材方面，突变是有重大意义的。所以大部分突变基因在种能够长期生存下来的环境中是有害的(detrimental)，但其中也有适合于不同环境的，在这个意义上，群体含有突变基因，对适应环境的变化，使种继續生存下去，可說是必要的条件。本章先假設十分大的群体，当其中新产生少数突变基因时，来看一看突变基因通过子孙是怎样地在群体中分布开去的。

§12 新产生的突变基因的固定和消失率

現在假定在相当大的群体中，产生一个突变，設这突变在次代

(以此为第一代)的群体中成为 k 个 $(k=0, 1, 2, \cdots)$ 的概率是 p_k。这突变再傳到下一世代时,假定各个突变基因相互独立,分布的法则和以前一样。于是設第一代的分布母函数为

$$\phi(x)=p_0+p_1x+p_2x^2+\cdots=\sum_{k=0}^{\infty}p_kx^k, \tag{12.1}$$

試来証明表示第二代分布的母函数 $\phi_2(x)$ 是

$$\phi_2(x)=\phi\{\phi(x)\}. \tag{12.2}$$

为此,暂且先假定有二个突变基因,我們来証明它們在次代的分布母函数为

$$\{\phi(x)\}^2.$$

事实上,突变基因数在次代成为 k 个的概率是

$$\sum_{i=0}^{k}p_ip_{k-i},$$

从而所求的母函数是

$$\sum_{k=0}^{\infty}\sum_{i=0}^{k}p_ip_{k-i}x^k=p_0^2+(p_0p_1+p_1p_0)x+(p_0p_2+p_1p_1+p_2p_0)x^2+\cdots$$

$$=(p_0+p_1x+p_2x^2+\cdots)^2=\{\phi(x)\}^2.$$

同样,一般有 k 个突变基因时,容易理解它們在次代的分布母函数为

$$\{\phi(x)\}^k.$$

因为从一个突变出发,在(次代,即)第一代,它成为特定数 k 个的概率是 p_k,由此,在第二代成为 n 个的概率是 $\{\phi(x)\}^k$ 中 x^n 的系数,所以知道,一般在第二代成为 n 个的概率是

$$\sum_{k=0}^{\infty}p_k\{\phi(x)\}^k$$

中 x^n 的系数。从 (12.1),这等于 $\phi\{\phi(x)\}$,从而証明了 (12.2)。在(12.2)的右边,如把 ϕ 看做算符,它对 x 施行二次,所以可写成 $\phi^2(x)$。一般,第 t 代 $(t\geqslant1)$ 的母函数可写成

$$\phi_t(x)=\phi\{\phi_{t-1}(x)\}=\phi_{t-1}\{\phi(x)\}=\phi^t(x). \quad (12.3)$$

又从以上的处理，如最初的突变基因数是 n 个，则可以知道，t 代后的分布母函数可写成

$$\{\phi_t(x)\}^n.$$

从(12.3)，很容易算出突变基因在后代的期望值(平均数)。現在設第 t 代的期望值为 M_t，則

$$M_t=\frac{d}{dt}[\phi_t(x)]_{x=1}=\phi'\{\phi_{t-1}(1)\}\phi'_{t-1}(1)$$
$$=\phi'(1)\phi'_{t-1}(1)=M_1M_{t-1},$$

所以

$$M_t=M_1^t. \quad (12.4)$$

同样，二次、三次等更高次的能率也可計算下去，特別在第 t 代的变差是

$$V_t=\phi''_t(1)+M_t(1-M_t), \quad (12.5)$$

但

$$\phi''_t(1)=\begin{cases}M_1^{t-1}(1-M_1^t)\phi''(1)/(1-M_1) & (M_1\neq 1\text{ 时}),\\ t\phi''(1) & (M_1=1\text{ 时}).\end{cases}$$

这里，$\phi''(1)=V_1-M_1(1-M_1)$。可是我們最感兴趣的是新生的突变基因的最后命运。对此，先証明下面的重要定理。

定理 12.1 新产生的一个突变，最后由群体中丧失的概率是方程

$$\phi(x)=x \quad (12.6)$$

的最小正根。

証明 由母函数的定义，这个突变在次代消失的概率是 $\phi(0)$，到第二代消失的概率是 $\phi^2(0)$，一般到第 t 代消失的概率是 $\phi^t(0)$。从(12.1)可以知道，$\phi(x)$ 是 x 的单調增加函数，若 $p_0\neq 0$，則

$$\phi(0)=p_0>0,$$

所以

$$\phi\{\phi(0)\}=\phi^2(0)>\phi(0),$$

一般

$$\phi^t(0)>\phi^{t-1}(0).$$

所以,如現在考虑数列

$$0, \phi(0), \phi^2(0), \phi^3(0), \cdots$$

这数列或是发散的，或具有某极限値 λ，但由概率母函数的性质可知它不可能是 ∞。所以这数列有极限値 λ，这 λ 就是所求的概率。对这样的 λ，

$$\lim_{t\to\infty}\phi^t(\lambda)=\lambda,$$

所以当然有

$$\phi(\lambda)=\lambda>0.$$

在数学上，λ 是 $\phi(x)$ 的固定点(fixed point)，对于 $0<x<\lambda$ 中的 x，数列

$$x, \phi(x), \phi^2(x), \phi^3(x), \cdots$$

在固定点是收斂的,所以知道，λ 必定是(12.6)的最小正根。

可是 $\phi(x)$ 与 $\phi'(x)$ 在区間 $0\leqslant x\leqslant 1$ 上都是单調增加函数，$\phi(0)=p_0>0$，$\phi(1)=1$，所以若

$$\phi'(1)=M_1\leqslant 1,$$

則 $\lambda=1$ 是满足(12.6)的最小正根。这个事实用作出图象来看是容易理解的。又若

$$\phi'(1)=M_1>1,$$

則满足(12.6)而比 1 小的正根只有一个。对应于 $M_1<1$，$M_1=1$ 和 $M_1>1$，突变基因在自然选擇上对野生型基因分别处于不利、中性和有利的关系，于是得到下面的重要結論。即新生的一个突变基因，只要它不是有利的，最后将从群体中消失，这几乎是肯定的。

为了进一步探討，必須設想(12.1)的母函数的具体型式。对于高等生物，特别是人类的群体，已經知道每一个体的孩子数大致上可以用 Poisson 分布来完满地描述，又为了数学上的簡化起見，先采用

$$\phi(x)=e^{-c}\left\{1+\frac{c}{1!}x+\frac{c^2}{2!}x^2+\cdots\right\}=e^{-c(1-x)} \qquad (12.7)$$

作为母函数是較方便的。在这个場合，$c(=M_1)$ 是次代的子孙基

因的平均数。因此，設突变基因到第 t 代为止从群体中消失的概率为 $p_0^{(t)}$，則

$$p_0^{(t)}=\phi_t(0),$$

所以由(12.3)和(12.7)，

$$p_0^{(t)}=e^{-c\left(1-p_0^{(t-1)}\right)}, \tag{12.8}$$

或設继續生存到第 t 代的概率为 $q_t(=1-p_0^{(t)})$，得

$$q_t=1-e^{-cq_{t-1}}. \tag{12.9}$$

由 $p_0=0$, $q_0=1$ 出发，用上式反复計算，可逐次算出在后代的生存率和消失率。某些場合的生存率揭示于表12.1。对于我們特別重要的，不消說是 t 非常大时的生存或消失的概率。先来看一看

表 12.1

世代数(t)	$c=0.90$	$c=0.95$	$c=0.97$	$c=0.99$	$c=1.00$
0	1.0000	1.0000	1.0000	1.0000	1.0000
1	0.5934	0.6132	0.6209	0.6284	0.6321
5	0.1972	0.2319	0.2462	0.2607	0.2680
10	0.0849	0.1186	0.1338	0.1499	0.1582
20	0.0236	0.0487	0.0625	0.0787	0.0875
50	0.0009	0.0076	0.0155	0.0287	0.0376
100	—	—	0.0028	0.0111	0.0193
200	—	—	—	0.0030	0.0098
300	—	—	—	0.0010	0.0065
400	—	—	—	—	0.0049
∞	0.0000	0.0000	0.0000	0.0000	0.0000

关于 $c=1$ 的中性基因。在这个場合，从定理12.1可以知道，

$$\lim_{t\to\infty} q_t=0,$$

所以当 t 大时，把(12.9)展开而得的

$$q_t=1-e^{-q_{t-1}}=1-\left(1-q_{t-1}+\frac{q_{t-1}^2}{2}-\cdots\right)=q_{t-1}-\frac{q_{t-1}^2}{2}+\cdots$$

可用微分方程

$$\frac{dq_t}{dt} = -\frac{1}{2}q_t^2$$

来代替，解这方程式，渐近地有

$$q_t \sim 2/t. \tag{12.10}$$

从(12.4)可以知道，突变基因的期望值每代是 1，所以在至少有一个这样基因的群体中，均值渐近于

$$t/2.$$

同样，也可求得变差及高次的能率，但实际上要求取子孙的突变基因数的分布，用矩母函数 (moment generating function) 较为便利。在现在的例子中，$c=1$，所以

$$\phi(x) = e^{x-1},$$

从而由(12.3)，得

$$\phi^t(x) = \phi^{t-1}(e^{x-1}).$$

所以如設第 t 代的频率分布的矩母函数为 $M^t(x)$，则利用

$$M^t(x) = \phi^t(e^x) = \sum_{n=0}^{\infty} \frac{\mu_n'^{(t)}}{n!} x^n,$$

得

$$M^t(x) = \phi^{t-1}(e^{e^x-1}) = M^{t-1}(e^x-1),$$

从而得

$$M^t(x) = M^{t-1}\left(\sum_{k=1}^{\infty} \frac{x^k}{k!}\right). \tag{12.11}$$

这里，$\mu_n'^{(t)}$ 是第 t 代分布的 0 附近的第 n 次能率。把(12.11)的两边展开，

$$\mu_0'^{(t)} + \frac{\mu_1'^{(t)}}{1!}x + \frac{\mu_2'^{(t)}}{2!}x^2 + \cdots + \frac{\mu_n'^{(t)}}{n!}x^n + \cdots$$

$$= \mu_0'^{(t-1)} + \frac{\mu_1'^{(t-1)}}{1!}x\left(1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots\right)$$

$$+ \frac{\mu_2'^{(t-1)}}{2!}x^2\left(1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots\right)^2 + \cdots$$

$$+ \frac{\mu_n'^{(t-1)}}{n!}x^n\left(1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots\right)^n + \cdots.$$

从而

$$\mu_1'^{(t)} = \mu_1'^{(t-1)},$$

$$\mu_2'^{(t)} = \mu_2'^{(t-1)} + \mu_1'^{(t-1)},$$

$$\mu_3'^{(t)} = \mu_3'^{(t-1)} + 3\mu_2'^{(t-1)} + \mu_1'^{(t-1)},$$

一般为

$$\mu_n'^{(t)} = \mu_n'^{(t-1)} + \frac{n(n-1)}{2}\mu_{n-1}'^{(t-1)} + \frac{n(n-1)(n-2)(3n-5)}{24}\mu_{n-2}'^{(t-1)} + \cdots. \tag{12.12}$$

这里若注意到 $\mu_n'^{(0)} = 1$，则反复应用上式，可渐次算出后代的分布能率。但从进化的观点来看，特别在 t 大时的分布是重要的。要求这分布，注意到 $\mu_n'^{(t)}$ 是与 t^{n-1} 同級的，故漸近地可置

$$\mu_n'^{(t)} = A_n t^{n-1}. \tag{12.13}$$

这里，A_n 是仅依賴于 n 而与 t 无关的常数。其次，用微分方程

$$\frac{d\mu_n'^{(t)}}{dt} = \frac{n(n-1)}{2}\mu_{n-1}'^{(t)}$$

来代替(12.12)，把(12.13)代入这微分方程中，得到

$$A_n = \frac{n}{2}A_{n-1},$$

从而如注意到 $A_1 = 1$，则得

$$A_n = n!/2^{n-1}.$$

所以(12.13)是

$$\mu_n'^{(t)} = n!\left(\frac{t}{2}\right)^{n-1}.$$

現在，設群体在第 t 代含有 k 个由突变而来的基因的概率为 $p_k^{(t)}$，则由(12.10)，得

$$\sum_{k=1}^{\infty} p_k^{(t)} \sim 2/t,$$

所以下面就来探討这种至少有一个突变基因的群体中的频率分布。設这分布的 n 次能率为 $\nu_n'^{(t)}$，則

$$\nu_n^{\prime(t)}=\mu_n^{\prime(t)}\Big/\left(\frac{2}{t}\right)=n!\left(\frac{t}{2}\right)^n,$$

从而矩母函数 $m(x)$ 是

$$m(x)=\sum_{n=0}^{\infty}\frac{\nu_n^{\prime(t)}}{n!}x^n=\sum_{n=0}^{\infty}\left(\frac{tx}{2}\right)^n=\frac{1}{1-\frac{tx}{2}}. \qquad (12.14)$$

已如上述，分布的均值是 $t/2$，仅在 t 大时成为問題，故应考慮把(12.14)近似地用連續分布来表示。若把 Gamma 分布的概率密度写作

$$f(X)=\frac{1}{\alpha!\beta^{\alpha+1}}X^{\alpha}e^{-X/\beta} \quad (X\geqslant 0),$$

則其矩母函数 $m(\theta)$ 为

$$m(\theta)=\int_0^{\infty}e^{X\theta}f(X)dX=\frac{1}{(1-\beta\theta)^{\alpha+1}},$$

所以如設 $\alpha=0$, $\beta=t/2$，則可使(12.14)近似于用概率密度

$$f(X)=\frac{2}{t}e^{-2X/t}$$

所求得的分布。或者把子孙突变基因数看作不超越 X 的頻率，得

$$F(X)=\int_0^X f(X)dX=1-e^{-2X/t}. \qquad (12.15)$$

Fisher(1930)把突变基因数作为 X 以上的頻率，而給出

$$e^{-2X/t},$$

就是基于上面所說的理由。現在假設发現有一个大的群体，其中存在着 5000 个突变率极低的中性突变基因。因为 $e^{-4.6}$ 約等于 0.01，所以可用 99% 以上的正确度来推論，这种突变基因的頻率，至少在过去 2000 年間，是由机会而漸次增加的。

在遺傳进化学上最有兴趣的情况是在群体中出現新的突变基因，在自然选擇上比原有的基因有利。現在設

$$c=1+s,$$

則 c 就是表示对自然选擇的有利程度 (selective advantage) 的尺度。对有利突变，不待説 $s>0$，要求这有利突变最后由群体中消失的概率，根据定理 12.1 和 (12.7)，可知这概率就是

$$x=e^{-(1+s)(1-x)}$$

的最小正根。因此，設在最后阶段的生存率为 q_∞，則

$$1-q_\infty=e^{-(1+s)q_\infty}. \tag{12.16}$$

这基因对野生型基因稍稍有利时，$0<s\ll 1$，所以 q_∞ 的值是小的，容易求得它的近似值。卽在 (12.16) 的两边取对数，得

$$-(1+s)q_\infty=\log(1-q_\infty)=-q_\infty-\frac{q_\infty^2}{2}-\cdots,$$

从而得到

$$q_\infty\approx 2s. \tag{12.17}$$

于第 t 代，在至少含有一个突变基因的群体中，根据 (12.4) 和 (12.5) 可知，突变基因的均值和变差分别近似地等于 $e^{st}/(2s)$ 和 $e^{2st}/(2s)$，因而当所取的世代数 t 十分大时，在这群体中突变基因数大于任意給定的大数 N 的概率，可任意地接近于 1。所以由 (12.17) 所得的生存率，就是群体中这个新突变确立的概率，也就是固定概率 (probability of fixation)。在再起突变 (recurrent mutation) 的場合，要上述的突变的固定概率大于 0.5，至少必須发生 $\log_e 2/(2s)$，卽 $(0.35)/s$ 次左右的突变。

这样，种在漫长年月間，位于恒定的环境中时，有利的再起突变几乎全部在群体中固定，除了极少产生的非再起突变中的极小一部分外，每代在群体中出現的突变，总是比野生型不利的。下面来考察一下群体中出現不利突变时，它們在后代中的命运。最近关于放射綫对人类群体的遺傳伤害問題，正作为社会問題被提到日程上来，所以在这个意义上，我們的探討也是重要的。在这个場合，(12.7) 中的 c 比 1 小，所以这种突变最后将由群体中消失。与

前同样，第 t 代的生存率 q_t 为

$$q_t = 1 - e^{-cq_{t-1}}, \qquad (12.18)$$

由 $q_0=1$，可逐次計算。与 $c=0.90, 0.95, 0.97, 0.99$ 相对应的生存率載于表 12.1。t 大时，渐近地成立

$$q_t = cq_{t-1},$$

所以到适当的世代 t_0，根据(12.18)来作数值計算，而后可由

$$q_{t+t_0} = c^t q_{t_0} \quad (q_{t_0} \ll 1) \qquad (12.19)$$

来求生存概率。关于群体中突变基因的频率分布，請参照卷末所列的作者論文(1956 b)。这里，来探討一下突变基因的平均寿命問題。某一突变基因生存到第 t 代，以后消失，其概率是 $q_t - q_{t+1}$，所以先把平均寿命定义为

$$\bar{t} = \sum_{t=0}^{\infty} t(q_t - q_{t+1}) = \sum_{t=1}^{\infty} q_t. \qquad (12.20)$$

給与 c 的值 $(0<c<1)$，要求出 $\bar{t}$ 时，到 q_{t_0} 成为很小为止，可用(12.18)計算，而后可用

$$\bar{t} = \sum_{1}^{t_0-1} q_t + \frac{q_{t_0}}{1-c} \qquad (12.21)$$

正确地算出所希望的位数。例如 $c=0.9$，则 $\bar{t}$ 約为 3.0。在上面的定义里，群体中生存的基因数未加考慮。若把随机的消失与增加的可能性平均起来，則各突变基因每代应以概率 c 継續生存，而以概率 $1-c$ 死灭，所以在这个意义上，应該把平均寿命定义为

$$\bar{\bar{t}} = \sum_{t=0}^{\infty} t(1-s)^{t-1}s = 1/s, \qquad (12.22)$$

但其中 $s=1-c$。为了跟 $\bar{t}$ 相区别，把 $\bar{\bar{t}}$ 叫做加权平均寿命。后者相当于 Muller (1950) 所說的突变持續性(persistence)。不待說，$\bar{\bar{t}}$ 比 $\bar{t}$ 大。中性基因 ($c=1$，从而 $s=0$) 的平均寿命和加权平均寿命，根据上面的定义来計算，都成为无限大：

$$\bar{t} = \bar{\bar{t}} = \infty.$$

但这样的基因是

$$\lim_{t\to\infty} q_t = 0,$$

所以最后将从群体中消失，几乎是肯定的。

在自然选擇上不利的基因，每代以頻率 v 在群体中再起时，按相对頻率

$$\hat{x} = v/s \tag{12.23}$$

达到平衡。把这关系看做是突变基因由于选擇而每代減少比率 s，而另一方面由于突变，每代以比率 v 从野生型基因新生，那就容易理解。不待說，这公式只有在 $v \ll s$ 的場合才能成立。但是根据人类群体方面的研究，关于有害性状的基因的 v，多在数量級 10^{-5} 以下，而另一方面对 s 的推算，平均是在 1/40 或以上（参照 Muller, 1955），所以 (12.23) 是很有利用价値的。把这公式与 (12.22) 組合，得

$$\hat{x} = v\bar{t}, \tag{12.24}$$

但这跟人类遺傳学上叫做 Danforth 的公式（参照 Muller, 1950）是相同的。

§13 基因頻率由于再起突变而生的变化

設基因 A 的頻率为 x，这基因由突变而变化为等位基因 A' 的頻率为 v。如沒有反方向的突变，則基因頻率每代的变化，用不連續模型，是

$$\Delta x = -vx,$$

从而

$$x_t = (1-v)^t x_0.$$

用連續模型，則

$$\frac{dx}{dt} = -vx,$$

从而得

$$x_t = x_0 e^{-vt}.$$

这里，$x_t(t=0,1,\cdots)$ 是第 t 代的 A 的頻率。以下为說明的便利起見，用連續模型。若有反方向的突变时，設其頻率为每代 u，則

$$\frac{dx}{dt}=-vx+u(1-x), \tag{13.1}$$

从而在点

$$\hat{x}=u/(u+v)$$

上达到平衡。(13.1) 又可写作

$$\frac{d}{dt}(x-\hat{x})=-(u+v)(x-\hat{x}),$$

根据这公式，知道跟平衡的偏差每代以頻率 $(u+v)$ 縮小。已如上述，突变率普通极低，而且几乎在大部分場合都导入了自然选擇，所以在自然群体中，不妨把单是由于突变而达到上述的平衡暂且看做是不存在的；因而探討自然选擇的效应，确实是必要的。現在，A 对 A' 的有利程度用 Malthus 参数来度量，假設是 $s(s>0)$，則如第 8 节习題 1 所示，x 由于选擇而起的变化率是

$$sx(1-x),$$

从而得

$$\frac{dx}{dt}=sx(1-x)-vx+u(1-x). \tag{13.2}$$

重要的情况是，u 和 v 比 s 小得多，这时大約在点

$$\hat{x}=1-(v/s) \tag{13.3}$$

上达到平衡。这跟由 (12.23) 所得的結果一致，平衡时的 A' 的頻率大約是 v/s。其次，来看一看 A 与 A' 間有显隐性关系的例子。先設 A' 为隐性，隐性表現个体 $A'A'$ 的适应度比显性表現个体即 AA 或 AA' 低 s，則如第 8 节习題 2 所示，x 仅由选擇而起的变化是 $sx(1-x)^2$，因而得

$$\frac{dx}{dt}=sx(1-x)^2-vx+u(1-x),$$

在 $u, v\ll s$ 的条件下，近似地得

$$\hat{x}=1-\sqrt{v/s}. \tag{13.4}$$

若 A' 方面是显性，$A'A'$ 或 AA' 的适应度用 Malthus 参数来度量时，比 AA 低 s，则

$$\frac{dx}{dt}=sx^2(1-x)-vx+u(1-x),$$

在类似条件下，

$$\hat{x}=1-v/s. \tag{13.5}$$

把这跟(13.4)比較，則隐性突变与显性突变的平衡频率之比是

$$\sqrt{v/s}:v/s,$$

前者要高得很多。換言之，显性突变的发生频率，即使跟隐性突变相同，但显性突变在群体中的出現頻率却要低得多。

§14 由突变而起的群体适应度的低下

根据上述的理由，現在群体中产生的变异，大部分是对个体不利的，所以可以設想，物种长期处于稳定的环境中，这些突变的蓄积，会使群体适应度低下。但不含遺傳性变异的群体，一旦遇到环境改变，在遺傳上不能进行相应的适应，就有絕灭的危險。在这意义上，由突变而起的适应度低下，好象是物种为了防备不測的灾难，而付了保險費一样。作为最簡单的例子，来看一看隐性突变，在平衡中的隐性个体 $A'A'$ 的频率是 $(\sqrt{v/s})^2$，即 v/s，由此，群体适应度的变化 $\Delta\bar{a}$ 是

$$(-s)(v/s)=-v,$$

即

$$\Delta\bar{a}=-v. \tag{14.1}$$

这里最重要的是，由突变而起的群体适应度的低下，不依存于各个突变基因的有害程度 $(-s)$，而仅由突变率所决定。大致地說，这是因为突变基因愈是有害，短期內就从群体消失，蓄积的程度不大，反之，突变基因的有害程度低，就被傳到后代，长期对群体发生微弱的危害。其次，突变是显性时，群体中显性表現个体的頻率是

$$2(v/s)(1-v/s)+(v/s)^2,$$

或略而为

$$2(v/s),$$

由此，群体适应度的变化是

$$(-s)(2v/s)=-2v,$$

即

$$\Delta\bar{a}=-2v. \qquad (14.2)$$

就一切有关的基因，把以上的結果合并起来，則起因于突变的群体适应度的变化是

$$\Delta\bar{a}=-\sum_i c_i v_i, \qquad (14.3)$$

其中 v_i 是第 i 个位点的突变率，c_i 是显性的程度，取 1 到 2 之間的値的常数。(14.3) 在討論放射能对人类群体的遺傳性危害上，是基础的公式。

习题 1 設群体內的近交系数是 F，$A\to A'$ 方向的突变率是 v，三遺傳型 AA，AA' 和 $A'A'$ 的适应度各为 0，$-h$，$-m$，則得

$$\Delta\bar{a}=-v-h(1-F)(1-\hat{x}),$$

从而証明 (a) 隐性有害突变，$h=0$，$\Delta\bar{a}=-v$，(b) 显性有害突变，$h=m$，$\Delta\bar{a}\simeq-v(2-F)$，(c) 不完全显性突变，

$$\Delta\bar{a}=-v\left\{1+\frac{h(1-F)}{mF+h-hF}-\frac{vh(1-F)^2(m-2h)}{(mF+h-hF)^3}+\cdots\right\}.$$

再，在 (c) 的場合，假定交配完全随机，則証明

$$\Delta\bar{a}\simeq-2v\left\{1-\frac{v(m-2h)}{2h^2}\right\},$$

只要 $2h^2\gg v(m-2h)$，結果在实质上跟完全显性的場合相等。

习题 2 考虑人类群体，設两亲在次代所留的平均子女数是 k，又每倫琴的每配子突变率为 μ。假定突变都是隐性，即使自然突变可以忽視，試証明为了人类群体的存續，每人一生的平均被照射量必須在

$$(\log_e k-\log_e 2)/\mu$$

倫琴以下。但設 $k>2$。[提示：設群体中交配完全随机，容許的倫琴量为 r 单位，則 $\bar{a}=\log(k/2)$，$\Delta\bar{a}=-\mu r$，从而由 $\bar{a}+\Delta\bar{a}\geqslant 0$，得 $\log(k/2)\geqslant\mu r$.]

第4章 群体遺傳学中的随机过程間題

§15 从随机过程看基因頻率的变化

从群体遺傳学的立場来看，进化中的最基本過程是群体的基因頻率的变化。最簡单的方法是把这个变化从决定論(deterministic)的角度来处理。若群体中含有非常多的个体，在漫長的岁月間，置于不变的环境中，則这个方法是可以完滿地运用的。

但在現实的群体中，由于各种因素的作用，基因頻率的变化不可能是決定論的。第一，在有性生殖中，伴随有配子的随机取样(random sampling of gametes)。基因頻率由此而起的随机变动，在小群体中特別显著。还有，这在决定各个突变基因对后代的命运，不論群体的大小，都有很大的关系。关于这个問題，在上一章的§12中已經討論过。第二，环境条件有不断的随机变动，这引起自然选擇的强度、移住率、突变率的随机变动。其中，对自然选擇强度的变动特別重要。

这样考虑起来，要在数学上探求基因頻率的变化，可以設想，会到达类似于統計力学那样的理論体系的。就作者所知，在这方面的研究以 Fisher 1922 年的論文为先驅。他在这篇論文中，处理了基因頻率的概率分布問題，就若干場合，求得定常状态的分布。虽然这篇論文有优美的想法，但数学的錯誤也不少，而且現在讀起来，也具有不能接受的結果。其后，經 Wright 指出錯誤，于1930年，Fisher 再就同样的問題訂正了以前的結果，使有进一步的发展。这篇論文恐怕是群体遺傳学中最美丽的論文之一。Wright 于 1931 年发表經典的論文“Mendel 群体的进化”以来，

继續这方面的研究，刊出了多数的論文。关于这些論文，請看卷末的参考文献。他集中力量，尽量以一般条件来解决定常状态的基因頻率的概率分布。他从这一研究得出关于进化机制的学說，但Fisher 一派与之正面对立，爭論不休。正如 Feller (1951) 所述，这两位学者用非常巧妙的方法求得了各个問題的解答。但是关于到达定常状态的过程几乎沒有研究。一方面在概率論的領域里，近年来随机过程的研究盛行，特别是 1931 年 Kolmogoroff 的論文发表以来，有着显著的发展。这一部分是因为，在自然科学的各个領域中，用随机过程理論来处理有关問題已一年年显得重要起来。因而在遺傳学領域中，想把基因頻率的時間变化在全过程中求得严密的解答，也是理所当然的。Malécot (1948) 談到突变与配子的随机取样时所用的求解方法，Goldberg (1950) 在未发表的博士論文中，成功地得到了它的解答。作者預先不知道有这些工作，开始研究，最近用各种条件，成功地得到了严密的解答。有兴趣的讀者請参照卷末所列的作者論文 (Kimura, 1954, 1955 a, b, c, 1956 a, b, 1957 b)。在群体遺傳学中所出現的拋物型偏微分方程，其普通边界是奇异点，不能課以任意的边界条件。Feller (1952) 用半群 (semi-group) 的理論，从純数学的立場，研究边界条件的問題，取得了优异的业績，但現在在群体遺傳学的具体問題的解答方面，还未能有直接的貢献。

上述的随机过程，特别是連續型随机过程的問題，近年来在群体遺傳学的数学理論上，愈益显得重要，但未解决的問題也不少。这里作为最簡单的一个例子，对所謂基因頻率的机会变动(random genetic drift)，略述其梗概。

§16　基因頻率的机会变动

假設有一隔离的群体，对于它，既沒有自然选擇，也沒有突变

和迁移，现在来看一看基因頻率每代由于配子的随机取样而起的变动。这个过程在这样单純化的形式上，引起很多生物学者的注意，給与各种的名称。Fisher (1922) 最初对它进行数学的研究，称之为"Hagedoorn effect"。从1931年 Wright 的研究以后，"drift"一詞在生物学者間极为流行，甚而出现"Wright drift"，"Sewall Wright 效应"等术語。但"drift"一詞本来在 Brown 运动等理論中指粒子的定向性运动，因此在"drift"之前如果不加形容詞"random"，就容易发生誤解，Wright 教授对此已再三强調指出。关于这方面研究的历史背景，請参照卷末所列的作者論文(Kimura, 1955 b)。

現在設群体的繁殖个体数为 N，考虑一对等位基因 A_1 和 A_2。在构成这群体的个体間，交配完全随机，所以次代的群体是由这一群体中随机取出的 N 个卵子与 N 个精子受精后构成的。設基因 A_1 的相对頻率为 x，A_2 的相对頻率为 $1-x$，則 x 是仅可取

$$0,\ \frac{1}{2N},\ \frac{2}{2N},\ \cdots,\ 1-\frac{1}{2N},\ 1$$

的值的不連續随机变数。在这种条件下继續繁殖下去，設每代 x 的变化为 δx，則 δx 服从均值为 0、方差为 $x(1-x)/(2N)$ 的二项分布，由 $x=i/(2N)$ 的群体，次代生 $x=j/(2N)$ 的群体，得到其轉移概率为

$$p_{j/i}=\binom{2N}{j}x^j(1-x)^{2N-j}\quad (x=i/(2N)).$$

所以这个过程是以

$$(p_{j/i})_{i,\,j=0,1,\cdots,2N}$$

为轉移矩陣的 Markoff 鏈。在这样的考虑下，Malécot (1944) 研究了这个过程，但不能得到满意的結果，特别是矩陣的特征值有誤，由 Feller (1951) 在以后加以訂正。不論如何，特征向量是复杂的，用这方法还不能得到完全的解答。但关于重要的分布的矩，已

得到表示随时間变化的公式(Robertson, 1952; Kimura, 1956 b)。

一般在自然群体中，构成这个群体的繁殖个体数 N 通常是相当大的，又在数量級 10^3 以下的个体数从进化上看来可以认为是暂时的，所以基因頻率 x 可近似地用連續随机变数来处理。在这場合，$0 \leqslant x \leqslant 1$，所以群体基因頻率的变化表現为实軸上 0 与 1 間的点的随机运动。先来考虑 $0<x<1$ 的 x。在遺傳学上，这表示群体中有二种等位基因共存的状态。設群体的最初基因頻率为 p，它在 t 代后成为 $x \sim x+dx$ 的概率密度为 $\phi(x|p;t)$。δx 的均值和方差各为 0 与 $x(1-x)/(2N)$，从連續随机过程的理論，ϕ 滿足下面的偏微分方程：

$$\frac{\partial \phi}{\partial t}=\frac{1}{4N} \frac{\partial^2}{\partial x^2}\{x(1-x)\phi\} \tag{16.1}$$

及初始条件

$$\phi(x|p;0)=\delta(x-p). \tag{16.2}$$

为了解(16.1)，設

$$\phi=T(t)X(x), \tag{16.3}$$

这里，假定 $T(t)$ 单是时間 t 的函数，$X(x)$ 单是基因頻率 x 的函数。把(16.3)代入(16.1)，两边用 $T \cdot X$ 除，则得

$$\frac{1}{T} \frac{\partial T}{\partial t}=\frac{1}{4NX} \frac{\partial^2}{\partial x^2}\{x(1-x)X\}. \tag{16.4}$$

但根据假定，这式的左边单是 t 的函数，右边单是 x 的函数，因而，左右两边的值必是常数。置这常数为 $-\lambda$，則由(16.4)得下面的二个常微分方程：

$$\frac{dT}{dt}=-\lambda T, \tag{16.5}$$

$$x(1-x)\frac{d^2X}{dx^2}+2(1-2x)\frac{dX}{dx}-(2-4N\lambda)X=0. \tag{16.6}$$

由(16.5)，除积分常数外，得到

$$T = e^{-\lambda t}. \tag{16.7}$$

其次，(16.6)相当于超几何微分方程

$$x(1-x)X'' + \{\gamma - (\alpha+\beta+1)x\}X' - \alpha\beta X = 0 \tag{16.8}$$

当 $\gamma=2$, $\alpha+\beta=3$, $\alpha\beta=2-4N\lambda$ 的場合。从而

$$\alpha = \frac{3+\sqrt{1+16N\lambda}}{2}, \quad \beta = \frac{3-\sqrt{1+16N\lambda}}{2}. \tag{16.9}$$

又，所給的偏微分方程(16.1)是以边界 $x=0$ 和 1 为奇点的，因而不能課以任意的边界条件。但是如下所示，求得了在边界上取有限値的解，即得到了問題的解答。容易了解，超几何微分方程(16.8)的二个独立解中，在現在的場合，仅 $F(\alpha, \beta, \gamma; x)$ 当 $x=0$ 时是有限的。其次，为了求得这个解在奇点 $x=1$ 成为有限的条件，可利用周知的关系

$$F(\alpha, \beta, 2, x) = \frac{\Gamma(2)\Gamma(2-\alpha-\beta)}{\Gamma(2-\alpha)\Gamma(2-\beta)} F(\alpha, \beta, -1+\alpha+\beta, 1-x)$$

$$+\frac{\Gamma(2)\Gamma(\alpha+\beta-2)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{2-\alpha-\beta}F(2-\alpha, 2-\beta, 3-\alpha-\beta, 1-x). \tag{16.10}$$

如注意到 $\alpha+\beta=3$，可知要使

$$\lim_{x\to1} F(\alpha, \beta, 2; x)$$

成为有限，$2-\alpha$ 必須为負的整数，又 β 必須为 0 或負的整数。所以可能的 λ 值为

$$\lambda = i(i+1)/(4N) \quad (i=1, 2, 3, \cdots). \tag{16.11}$$

对这特征値，$\alpha=2+i$, $\beta=1-i$，因而除常数因子外，得到

$$X = F(2+i, 1-i, 2; x). \tag{16.12}$$

或应用 Gegenbauer 多項式

$$T_{i-1}(z) = \frac{i(i+1)}{2} F\left(i+2, 1-i, 2; \frac{1-z}{2}\right), \tag{16.13}$$

除常数因子外，得到

$$X=T_{i-1}(z),$$

但設 $z=1-2x$。这里，

$$T_0(z)=1,\ T_1(z)=3z,\ T_2(z)=\frac{3}{2}(5z^2-1),$$

$$T_3(z)=\frac{5}{2}(7z^3-3z),\ \cdots.$$

合并以上結果，得到所求的解的形式为

$$\phi(x|p;t)=\sum_{i=1}^{\infty}C_iT_{i-1}(z)e^{-i(i+1)t/(4N)}. \tag{16.14}$$

这里，C_i 是常数，可用初始条件(16.2)决定。 即对

$$\sum_{i=1}^{\infty}C_iT_{i-1}(z)=\delta(x-p), \tag{16.15}$$

应用正交关系

$$\int_{-1}^{1}(1-z^2)T_m(z)T_{i-1}(z)dz=\delta_{m,i-1}\frac{2(i+1)i}{(2i+1)},$$

可得

$$C_i=4pq\frac{(2i+1)}{i(i+1)}T_{i-1}(1-2p).$$

因而所求的解是

$$\phi(x|p;t)=\sum_{i=1}^{\infty}\frac{(2i+1)(1-r^2)}{i(i+1)}T_{i-1}(r)T_{i-1}(z)e^{-i(i+1)t/(4N)}$$
$$=6pqe^{-t/(2N)}+30pq(q-p)(1-2x)e^{-3t/(2N)}+\cdots. \tag{16.16}$$

但設 $r=1-2p$, $q=1-p$。右边的級数对于 $t>0$ 是一致收斂的。

設群体中两等位基因 A_1 和 A_2 共存的概率为 Ω_t，則

$$\Omega_t=\int_0^1\phi(x|p;t)dx$$
$$=\sum_{m=1}^{\infty}\frac{(4m-1)(1-r^2)}{(2m-1)2m}T_{2m-2}(r)e^{-(2m-1)2mt/(4N)}$$
$$=6pqe^{-t/(2N)}+14pq(-5pq+1)e^{-6t/(2N)}+O(e^{-15t/(2N)}). \tag{16.17}$$

与上面一样，右边的級数对于 $t>0$ 是一致收敛的。再，也可以証明

$$\lim_{t\to 0}\Omega_t=1;\quad \lim_{t\to\infty}\Omega_t=0.$$

这样，群体中两等位基因的共存概率随时間而减少，終于成为 0。这跟基因的固定和消失在现在所考虑的条件下是不可逆的这种遺傳上的直覚相一致。在这种情况下，边界 $x=0$ 和 $x=1$ 的作用应該是所謂“吸收壁”(absorbing barrier)的作用，其两端所吸收的概率的量就是基因的固定和消失的概率。为計算这概率，設基因 A_1 到第 t 代在群体中固定的概率为 $f(1|p;t)$，完全消失的概率(A_2 固定的概率)为 $f(0|p;t)$，則

$$\begin{aligned} df(1|p;t)/dt&=\phi(1|p;t)/(4N),\\ df(0|p;t)/dt&=\phi(0|p;t)/(4N). \end{aligned}\tag{16.18}$$

关于上式的証明，因篇幅关系从略。只把应用这些公式后所得的結果列出如下：

$$\begin{aligned} f(1|p;t)&=\int_0^t\phi(1|p;t)dt/(4N)\\ &=p+\sum_{i=1}^{\infty}\frac{(-1)^i}{2}\{P_{i-1}(r)-P_{i+1}(r)\}. \end{aligned}\tag{16.19}$$

这里 $r=1-2p$, $P_n(r)$ 表示 Legendre 多項式。消失概率可从

$$f(0|p;t)=f(1|q;t)\tag{16.20}$$

求得。于是，把已求得的三个概率(16.16)，(16.19)，(16.20)加起来，則如所預期，可以証明是 1：

$$f(1|p;t)+\Omega_t+f(0|p;t)=1.$$

設由群体中随意取出的个体为异型接合体 A_1A_2 的概率为 H_t，則得

$$\begin{aligned} H_t&=\int_0^1 2x(1-x)\phi(x|p;t)dx\\ &=2pqe^{-t/(2N)}. \end{aligned}\tag{16.21}$$

即在群体中，异型接合体的頻率每代以 $1/(2N)$ 的比率降低。

以上是群体中含有二个等位基因时的理論，若含有三个以上的复等位基因，則計算极为复杂。例如在三复等位基因时，設这三基因为 A_1, A_2, A_3, 在时間 t 的頻率为 x_1, x_2, $1-x_1-x_2$, 又初始頻率为 p_1, p_2, $1-p_1-p_2$, 則所求的轉移概率密度 $\phi(x_1, x_2|p_1, p_2; t)$ 滿足下面的偏微分方程：

$$\frac{\partial\phi}{\partial t}=\frac{1}{4N}\frac{\partial^2}{\partial x_1^2}\{x_1(1-x_1)\phi\}-\frac{1}{2N}\frac{\partial^2}{\partial x_1\partial x_2}\{x_1x_2\phi\}+\frac{1}{4N}\frac{\partial^2}{\partial x_2^2}\{x_2(1-x_2)\phi\}\quad(0<x_1<x_1+x_2<1).$$

这时的初始条件是

$$\phi(x_1, x_2|p_1, p_2; 0)=\delta(x_1-p_1)\delta(x_2-p_2).$$

一般含有 n 个复等位基因 A_1, A_2, $\cdots$, A_n 时，也可成立同样的偏微分方程，但到現在为止，其严密的解答还无法求得。

参考文献

[1] J. F. Crow: Cold Spring Harbor Symp. Quant. Biol., **20** (1955), 54.

[2] J. F. Crow and M. Kimura: Proc. Third Berkeley Symp. Math. Stat. and Probability, **4** (1956), 1.

[3] W. Feller: Proc. Second Berkeley Symp. Math. Stat. and Probability, **2** (1951), 227.

[4] W. Feller: Ann. Math., **55** (1952), 468.

[5] R. A. Fisher: Trans. Roy. Soc., **52** (1918), 399.

[6] R. A. Fisher: Proc. Roy. Soc. Edinb., **42** (1922), 321.

[7] R. A. Fisher: Proc. Roy. Soc. Edinb., **50** (1930), 205.

[8] R. A. Fisher: Ann. Eugenics, **11** (1941), 53.

[9] S. Goldberg: Ph. D. Thesis (Cornell Univ.) (1950).

[10] M. Kimura: Genetics, **39** (1954), 280.

[11] M. Kimura: Proc. Nat. Acad. Sci., **41** (1955 a), 144.

[12] M. Kimura: Evolution, **9** (1955 b), 419.

[13] M. Kimura: Cold Spring Harbor Symp. Quant. Biol., **20** (1955 c), 33.

[14] M. Kimura: Biometrics, **12** (1956 a), 57.

[15] M. Kimura: Ph. D. Thesis (Univ. of Wisconsin) (1956 b).
[16] M. Kimura: Proc. Nat. Acad. Sci., **42** (1956 c), 336.
[17] M. Kimura: Evolution, **10** (1956 d), [illegible].
[18] M. Kimura: Proc. International Genetics Symposia (1957 a) (in press).
[19] M. Kimura: Ann. Math. Stat. (1957 b) (in press).
[20] A. Kolmogoroff: Math. Ann., **104** (1931), 415.
[21] G. Malécot: Compt. Rend. de l'Acad. des Sciences, **219** (1944), 379.
[22] H. J. Muller: Amer. Jour. Human Genetics, **2** (1950), 111.
[23] H. J. Muller: Bull. Atomic Scientists, **9** (1955), 329.
[23] A. Robertson: Genetics, **37** (1952), 189.
[24] C. Stern: Science, **97** (1943), 137.
[25] S. Wright: Genetics, **6** (1921), 111, 124, 144, 162, 167.
[26] S. Wright: Genetics, **16** (1931), 97.
[27] S. Wright: Jour. Genetics, **30** (1935), 343.
[28] S. Wright: Ann. Eugenics, **15** (1951), 323.
[29] S. Wright: Cold Spring Harbor Symp. Quant. Biol., **20** (1955), 16.
[30] R. A. Fisher: The Genetical Theory of Natural Selection (Clarendon Press, Oxford, 1930).
[31] O. Kempthorne: Introduction to Genetic Statistics (Wiley, New York, 1957).
[32] I. M. Lerner: Genetic Homeostasis (Wiley, New York, 1954).
[33] G. Malécot: Les Mathématiques de l'Hérédité (Masson, Paris, 1948).
[35] K. Mather: Biometrical Genetics (Methuen, London, 1949).
[36] A. H. Sturtevant and G. W. Beadle: An Introduction to Genetics (Saunders, Philadelphia).